

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 00:52:06 ; Search time 73.54 Seconds
(without alignments)
85.089 Million cell updates/sec

Title: US-09-589-870-2
Perfect score: 183
Sequence: 1 MRKIVAAIVASLTTSITL.....IDAAKAGVNGNPLDAVQ 183

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 268485 seqs, 34193795 residues

Word size : 0
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_36:*

1: /cgn2_2/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /cgn2_2/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /cgn2_2/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /cgn2_2/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /cgn2_2/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /cgn2_2/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /cgn2_2/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /cgn2_2/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /cgn2_2/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /cgn2_2/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /cgn2_2/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /cgn2_2/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /cgn2_2/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /cgn2_2/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	183	7	P60625
2	183	100.0	183	10	P93530
3	183	100.0	183	14	R44491
4	183	100.0	183	18	W29306
5	183	100.0	183	19	W59216
6	183	100.0	183	20	V17868
7	183	100.0	183	21	V44701
8	183	100.0	186	21	V60512
9	182	99.5	184	21	V80513
10	160	87.4	161	21	V80515
11	160	87.4	435	15	R56483
12	159	86.9	160	10	P93531

13	159	86.9	162	21	V80514	Streptomycetes avidi
14	159	86.9	163	9	P80160	Biosynthetic prote
15	152	83.1	182	8	P70492	Streptavidin sequ
16	149	81.4	183	21	V84020	Amino acid sequenc
17	140	76.5	415	15	R56484	Scfv PRAS109 and p
18	134	73.2	159	18	W29309	Streptavidin prote
19	132	72.1	159	18	W29310	Streptavidin prote
20	128	69.9	168	21	V44700	Potato proteinase
21	127	69.4	128	14	R34722	Core streptavidin.
22	127	69.4	128	17	W04211	Streptomycetes avidi
23	127	69.4	128	18	W29308	Recombinant Core-s
24	127	69.4	270	20	V28928	S. avidinii protein
25	127	69.4	673	17	W04208	Streptavidin/Lucif
26	123	67.2	140	17	W04210	Mutant streptomyc
27	123	67.2	685	17	W04209	Streptavidin
28	121	66.1	122	21	V80516	Streptomycetes avidi
29	121	66.1	122	21	V80517	Streptavidin prote
30	119	65.0	159	18	W29312	Streptavidin prote
31	114	62.3	159	18	W29311	Streptavidin prote
32	114	62.3	159	18	W29313	Streptavidin prote
33	114	62.3	159	18	W29314	Streptavidin prote
34	112	61.2	183	19	W59217	S. avidinii strept
35	112	61.2	183	19	W59218	S. avidinii strept
36	88	48.1	186	21	V84021	Amino acid sequenc
37	88	48.1	186	21	V84022	Amino acid sequenc
38	74	40.4	159	18	W29320	Streptavidin prote
39	69	37.7	402	15	R56485	Scfv PRAS110 and p
40	64	35.0	159	18	W29315	Streptavidin prote
41	64	35.0	159	18	W29316	Streptavidin prote
42	64	35.0	159	18	W29317	Streptavidin prote
43	64	35.0	159	18	W29318	Streptavidin prote
44	64	35.0	159	18	W29319	Streptavidin prote
45	40	21.9	40	8	P70491	N-terminal sequenc

ALIGNMENTS

RESULT	1
P60625	
ID	P60625 standard; Protein; 183 AA.
XX	
AC	P60625;
XX	
DT	13-AUG-1991 (first entry)
XX	
DE	Sequence of a streptavidin-like polypeptide encoded by SA307.
XX	
KW	Antibiotic; biotin binding affinity; fusion protein.
XX	
OS	Streptomycetes.
XX	
PN	W08602077-A.
XX	
PD	10-APR-1986.
XX	
PF	01-OCT-1985; 85WO-0001901.
XX	
PR	02-OCT-1984; 84US-0656873.
XX	
PA	(MEAD/) MEADE H M.
XX	
PI	Meade HM, Garvin JL, Biogen NV;
XX	
DR	WPI: 1986-106643/16.
XX	
XX	N-PSDB; N60626.
PT	DNA sequences and hybrid DNA sequences - encoding
XX	streptavidin-like polypeptide, also joined to another protein,
XX	e.g. tissue plasminogen activator
PS	Disclosure; Fig. 2; 54pp; English.
XX	

CC The inventors claim the DNA sequence in SA307 which codes for a
 CC streptavidin-like polypeptide (see M60626), and the polypeptide
 CC encoded by it (P60625). They also claim hybrid SQs comprising M60626
 CC and a second sequence coding for another protein, polypeptide,
 CC peptide or AA (pref. tissue plasminogen activator (TPA)).

SQ Sequence 183 AA;

Query Match 100.0%; Score 183; DB 7; Length 183;
 Best Local Similarity 100.0%; Pred. No. 7.6e-173;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIYVAALAVSLTIVTSINASADPSKSKAQVSAEAGITGTWYNQGSTFIYTAGAD 60
 |
 Db 1 mkriyvaalavslitvstiasasadpskkskagsaaegitgtwynqgstfiytagad 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120
 |
 Db 61 galtgtyesavgnaesryvltgrysapatsdgsctalgwtvwmknnyrnahsattwsqgy 120
 QY 121 VCGAERINTOMLITSGTEANAMKSTLVGHDTFTVKRPSAASIDAKKAGVNNGNPLDA 180
 |
 Db 121 vpgaearintqwlitsgteanawksclvghdtftkvkpsaasidaakkagvnngnplda 180
 QY 181 VQO 183
 |||
 Db 181 vqg 183

RESULT 2
 ID P93530 standard; protein; 183 AA.
 AC P93530;

DT 04-JUN-1990 (first entry)

XX Streptavidin protein.

XX Streptavidin; Streptomyces avidinii; biotin.

XX Streptomyces avidinii.

OS Streptomyces avidinii.

XX Key Location/Qualifiers

FT Peptide 1..24

FT Protein /note="Leader sequence"
 25..159
 /note="this sequence was as the basis for the design
 for the synthetic gene of the present invention."

XX W08903422-A.

XX 20-APR-1989.

XX 07-OCT-1988; 88WO-GB00831.

XX 08-OCT-1987; 87GB-0023661.

XX (BRBI-) BRIT BIO-TECHN LTD.

XX Edwards RM;

XX WPI; 1989-130040/17.

XX DNA sequence encoding streptavidin and vector -
 PT comprising hybrid gene encoding fusion protein with
 PT biotin-binding activity

PS Fig 1; page 1/5; 22pp; English.

CC Streptavidin is a 60KD protein isolated from Streptomyces avidinii that
 CC binds extremely tightly to the vitamin biotin. It is composed of four

CC identical subunits of 15KD and binds 4 mole of biotin per mole of
 CC protein. It is structurally related to the protein avidin. It can be
 CC readily conjugated to a range of other proteins. In order to facilitate
 CC the incorporation of streptavidin into expression vectors and the
 CC production of novel chimeric proteins containing streptavidin
 CC functionality, an improved novel synthetic gene for streptavidin has
 CC been constructed (n90755) based on the amino acid sequence of mature
 CC streptavidin.

SQ Sequence 183 AA;

Query Match 100.0%; Score 183; DB 10; Length 183;
 Best Local Similarity 100.0%; Pred. No. 7.6e-173;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIYVAALAVSLTIVTSINASADPSKSKAQVSAEAGITGTWYNQGSTFIYTAGAD 60
 |
 Db 1 mkriyvaalavslitvstiasasadpskkskagsaaegitgtwynqgstfiytagad 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120
 |
 Db 61 galtgtyesavgnaesryvltgrysapatsdgsctalgwtvwmknnyrnahsattwsqgy 120
 QY 121 VCGAERINTOMLITSGTEANAMKSTLVGHDTFTVKRPSAASIDAKKAGVNNGNPLDA 180
 |
 Db 121 vpgaearintqwlitsgteanawksclvghdtftkvkpsaasidaakkagvnngnplda 180
 QY 181 VQO 183
 |||
 Db 181 vqg 183

RESULT 3
 ID R44491 standard; protein; 183 AA.
 AC R44491;

DT 27-JUN-1994 (first entry)

XX Streptavidin gene.

XX Streptavidin; protein secretion; Bacillus subtilis.

XX Streptomyces avidinii.

OS Streptomyces avidinii.

XX Key Location/Qualifiers

FT Protein /label= signal_peptide

FT Peptide 1..183

FT /label= streptavidin

XX W09324631-A.

XX 09-DEC-1993.

XX 27-MAY-1993; 93WO-US05240.

XX 29-MAY-1992; 92US-0891524.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Nagarsajan V;

XX WPI; 1993-405822/50.

XX P-PSDB; Q53412.

PT Streptavidin prodn. from Bacillus subtilis - using signal protein
 PT from bacterial exo-protein and expression element from Gram

PT positive bacterial protein.
XX
PS Disclosure: Fig 1b; 54pp; English.
XX
CC Tetrameric biologically active streptavidin is produced by secretion
CC from *Bacillus subtilis* transformed with a plasmid encoding the
CC sequence.
XX
SQ Sequence 183 AA:

Query Match 100.0%; Score 183; DB 14; Length 183;
Best Local Similarity 100.0%; Pred. No. 7.6e-173;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIIVAAIAVSLITVSTIRASADPSKDSKAQVSAEAGITWTWNLGSTRFTVAGAD 60
DB 1 mrkivvaalavslitvstirasaadpskdskaqvsaagltgwtwnlgstrftvtagad 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVMKNNYRNAHSATWMSGOY 120
DB 61 galtgtyesavgnaesryvltgrydsapatdgsgtalgtvwmknnyrnhsatwsggy 120
QY 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDPFTYKPPSAASIDAAKKAGVNNGNPLDA 180
DB 121 vggaeairintqwlitsgtteanawksltvghdftkyvkpsaasidaakkagvnngnplda 180
QY 181 VQO 183
DB 181 vqg 183

RESULT 4
ID W29306 standard; Protein: 183 AA.
XX
AC W29306;
XX
DT 27-APR-1998 (first entry)
XX
DE Wild-type streptavidin protein.
XX
DE Streptavidin: biotin; anti-interference reagent; detection; mutein;
KM avidin; non-specific binding.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Peptide 1..24
FT Protein /label= signal
FT Protein 25..183
XX
PN DE19637718-A1.
XX
PD 02-OCT-1997.
XX
PE 16-SEP-1996; 96DE-1037718.
XX
PR 01-APR-1996; 96DE-1013053.
XX
PA (BOE) BOEHRINGER MANNHEIM GMBH.
XX
PI Brandstetter H, Deger A, Engn R, Kopeckzi E, Mueller R;
PI Schmitt U;
XX
DR WPI: 1997-482043/45.
DR N-PSDB: T73193.
XX
PT Streptavidin and avidin muteins with reduced binding affinity for
PT biotin - useful for reducing interference from nonspecific binding
PT in assays
XX
PS Disclosure: Page 17-18; 26pp; German.

XX
CC This sequence represents a streptavidin which is used in a novel method
CC of reducing interference from non-specific binding in assays. Muteins
CC constructed from a core streptavidin or avidin sequence are selected that
CC differ from the native polypeptide by at least one amino acid and have a
CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable
CC polypeptide may be present as a polymeric conjugate, e.g. with another
CC polypeptide or protein, especially bovine serum albumin. These muteins
CC are used as anti-interference reagents for reducing and/or avoiding
CC nonspecific interactions in a process for detecting an analyte. In
CC particular, they are used in assays where the streptavidin/avidin-biotin
CC specific binding pair is involved for qualitative and/or quantitative
CC determination of an analyte in a test sample, e.g. a heterogeneous
CC immunoassay or a hybridisation assay. Despite having a lower binding
CC affinity for biotin, the muteins have high immunological cross-reactivity
CC with native streptavidin and avidin.
XX
SQ Sequence 183 AA:

Query Match 100.0%; Score 183; DB 18; Length 183;
Best Local Similarity 100.0%; Pred. No. 7.6e-173;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIIVAAIAVSLITVSTIRASADPSKDSKAQVSAEAGITWTWNLGSTRFTVAGAD 60
DB 1 mrkivvaalavslitvstirasaadpskdskaqvsaagltgwtwnlgstrftvtagad 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVMKNNYRNAHSATWMSGOY 120
DB 61 galtgtyesavgnaesryvltgrydsapatdgsgtalgtvwmknnyrnhsatwsggy 120
QY 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDPFTYKPPSAASIDAAKKAGVNNGNPLDA 180
DB 121 vggaeairintqwlitsgtteanawksltvghdftkyvkpsaasidaakkagvnngnplda 180
QY 181 VQO 183
DB 181 vqg 183

RESULT 5
ID W59216 standard; Protein: 183 AA.
XX
AC W59216;
XX
DT 27-AUG-1998 (first entry)
XX
DE S. avidinii streptavidin protein.
XX
DE Streptavidin: ligand; binding affinity; mutant; isolation;
KM purification; recover; immobilise.
XX
OS Streptomyces avidinii.
XX
PN EP835934-A2.
XX
PD 15-APR-1998.
XX
PE 09-OCT-1997; 97EP-0117504.
XX
PR 10-OCT-1996; 96DE-1041876.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Skerra A, Voss S;
XX
DR WPI: 1998-218668/20.
DR N-PSDB: V34714.
XX
PT Streptavidin mutants with higher binding affinity for peptide
PT ligands - have mutation in amino acid region 44-53, used to isolate,
XX

PT purify or determine fusion proteins including these ligands
XX
PS Disclosure; Page -; 21pp; German.
v

CC This sequence encodes a wild-type streptavidin protein isolated from
CC streptomycetes avidinii. This sequence is used to produce mutants which
CC are used in a method to assay the binding affinity of streptavidin
CC mutants. These mutants have a mutation within the amino acid (aa) region
CC 44-53 of the wild-type protein show a higher binding affinity than the
CC wild-type for peptide ligands that include the sequence of formula
CC $\text{Trp-X-His-Pro-Gln-Phe-Y-Z}$ where X = any aa; Y and Z are both gly,
CC or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can
CC be used to isolate, purify and determine proteins or to determine/recover
CC substances that contain streptavidin-binding groups. Such compounds may
CC also be used to immobilise fusions on microtitre plates, microbeads or
CC sensor chips.
CC NOTE: This sequence does not appear in the specification but is used to
CC make the mutant streptavidin proteins represented in W59217 and W59218.
XX
Q0 Sequence 183 AA;

Query Match	100.0%	Score 183;	DB 19;	Length 183;
Best Local Similarity	100.0%	Pred. No. 7.6e-173;		
Matches 183; Conservative	0;	Mismatches	0;	Indels 0

[illegible]

RESULT	6
Y17868	
ID	Y17868 standard; Protein; 183 AA.

DT	20-AUG-1999 (first entry)
XX	
DE	Streptococcus streptavidin.

KW Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;
hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin

OS	Streptococcus sp.
XX	
PN	W09929838-A1.

PD 17-JUN-1999.

PF 09-DEC-1998; 98WO-US26086.

PR 09-DEC-1997; 97US-0067978.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

Cederholm-Williams SA;

DR WPI; 1999-385599/32.
DR N-PSDB; X80198.

PT A fibrinogen-converting enzyme fusion protein
XX
PS Disclosure; Page 28; 35pp; English.

The present invention describes a fibrinogen-converting enzyme fusion protein (FCE). The fusion protein is a multidomain protein comprising: (a) a FCE; and (b) a first member of a binding pair, that is linked to the FCE chain: (i) directly by bonds utilizing the N-terminal amino groups, the C-terminal carboxy groups or side-chain functionalities; (ii) via a bifunctional linkage moiety linking the groups or functionalities; or (iii) by the first member binding to the second member of the binding pair, where the second member of the binding pair is covalently attached to the first polypeptide chain. The FCE can be used in a method for producing fibrin. Fibrin is useful as a sealant in surgery to, e.g., reduce bleeding by sealing blood vessels, and tissues that have been dissected either in surgery or through wounding. The fusion protein allows for the removal of the fibrinogen converting enzyme from the fibrin sealant preparation via the binding of streptavidin to a biotin solid support. The present sequence represents *Streptococcus streptavidin* as given in the present invention.

Query March	100.0%;	Score 183;	DB 20;	Length 183;
Best Local Similarity	100.0%;	Pred. No. 7.6e-173;		
Matches 183; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	MRIKIVAAIVSLTYSITASASDPERKSDKAQVSAEADITGWTYQOLSTFIVTGAD	60
Db	1	mrkivaaivsltvtasasdpkdskaqysaaegltgwyqlgstfivtagad	60
QY	61	GALITGETEASVGNNEERYVLTRGRDSPAIDDGSTALGTVAAKNKNRNHSAFTMSGQY	120
Db	61	galitgetesavgnaesryvltgrdsapadcdgstaqlgvtwknmyrnahsatltsqgy	120
QY	121	VGGAEARINIQWLITSGTTEANAMKSTLVGHDFITKYKPSAASIDAAKKAGVNNGNPLDA	180
Db	121	vggaeaarinqwlitstteanawksltvghdftkykpsaasidaakkagvnnnpida	180

RESULT	7
Y44701	
ID	Y44701 standard; Protein; 183 AA

DT 25-APR-2000 (first entry)

Streptavidin protein for recombinant pART27 vector.

KM potato protease inhibitor-II, ppi-II, streptavidin; worm;
 KM insect, plant-noxious protein, pest resistance; moth; insect; weevil;
 KM grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;
 insecticidal.

Unidentified OS

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```
FT /note= "Signal peptide"
```

PN WO200004049-A1

PD 27-JAN-2000.

PF 15-JUL-1999; 99WO-NZ00110

PR 15-JUL-1998; 98NZ-0331002


```
XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
XX Christeller JF, Sutherland PW, Murray C, Markwick NP, Phillip BA,
XX Malone LA, Burgess EPJ;
XX WPI: 2000-171244/15.
XX DR N-PSDB: 249867.
XX
XX New chimeric polypeptide and composition comprising the polypeptide
XX useful for conferring pest resistance on plants
XX
XX Disclosure: Fig 12; 11pp; English.
XX
XX The present sequence is streptavidin, a plant-toxicous protein.
XX Recombinant vector, pAK27 expressing a chimeric polypeptide comprising
XX streptavidin mature peptide fused to the potato proteinase inhibitor-II
XX (PPI-II) signal peptide is targeted to the vacuole.
XX Transformation of plant genome with the vector can produce pest
XX resistance in plants, plant derived products and stored harvest
XX material. Pests that can be controlled include, cotton bollworm,
XX tropical army worm, European corn-borer or red mite, tobacco horn worm,
XX loopers, rice stem borer, portia, cutworms, diamondback moth, potato
XX tuber moth, codling moth, indian meal moth, gypsy moth, argentine stem
XX weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat
XX weevils, mealworms, flour beetles, black field cricket, locusts,
XX sawflies, Western flower thrips, Hessian flies or two-spotted mite.
XX
XX Sequence 183 AA:
SQ
Query Match 100.0%; Score 183; DB 21; Length 183;
Best Local Similarity 100.0%; Pred. No. 7.6e-173;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRIIVAAIAVSLTIVSITRNASADPSKDSKAQVSAEAGITGTWYNOLGSTRFYTAGAD 60
DB 1 mKRIIVAAIAVSLTIVSITRNASADPSKDSKAQVSAEAGITGTWYNOLGSTRFYTAGAD 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDAPATDGSCTALGWTVMKNNYRNAHSAATTWSGQY 120
DB 61 gAltgtYesaVgnaesryvltgrYdsapAtdgsctAlgwtVmkNnyrNahsaAttwsGqy 120
QY 121 VGGAERINTOMLITSGTEBANAMKSTLVGHDTFTVKRPSAASIDAKKAGVNGNPIDA 180
DB 121 vGgaerIntomLiTsGteBanamKstlvGHdtftVkrPsAaSIdaKkAgvNgnpIda 180
QY 181 VQG 183
DB 181 vqg 183
RESULT 8
Y80512
ID Y80512 standard; Protein: 186 AA.
XX
XX Y80512;
XX
XX 06-JUN-2000 (first entry)
XX
XX Streptomyces avidinII sps protein.
XX
XX Plant somatic tissue degeneration; plant essential factor; depletion;
XX viability; sps gene; plant development; plant morphology; flower;
XX fruit plant.
XX
XX Streptomyces avidinII.
XX
XX WO200007427-A2.
XX
XX 17-FEB-2000.
XX
XX 30-JUL-1999; 99WO-IL00420.
XX
```

```
XX
XX 03-AUG-1998; 98IL-0125632.
XX
XX (AGRI-) AGRIC RES ORG.
XX
XX Kapulnik Y, Ginzberg I;
XX WPI: 2000-195402/17.
XX DR N-PSDB: 291073.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
XX protein, useful for controlling plant development and morphology, such
XX as decreasing the number of flowers present to increase the number of
XX fruit -
XX
XX Examples: Page 84; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
XX plant tissue by expressing a heterologous protein capable of binding a
XX plant essential factor (PEF), in somatic plant tissue cells, where
XX heterologous protein expression causes depletion of the PEF so the plant
XX viability is maintained, while simultaneous degeneration of the somatic
XX plant tissue is effected. This sequence represents the Streptomyces
XX avidinII sps protein as an example of a heterologous protein introduced
XX into the plants. The methods can provide for the selective and optionally
XX reversible cell degeneration in somatic plant tissue. They can be used
XX for artificially controlling plant development and morphology. They can
XX be used e.g. to decrease the number of flowers in fruit producing plants
XX so as to increase the number of fruits which reach maturity.
XX
XX Sequence 186 AA:
SQ
Query Match 100.0%; Score 183; DB 21; Length 186;
Best Local Similarity 100.0%; Pred. No. 7.7e-173;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRIIVAAIAVSLTIVSITRNASADPSKDSKAQVSAEAGITGTWYNOLGSTRFYTAGAD 60
DB 4 mKRIIVAAIAVSLTIVSITRNASADPSKDSKAQVSAEAGITGTWYNOLGSTRFYTAGAD 63
QY 61 GALTGTYESAVGNAESRYVLTGRYDAPATDGSCTALGWTVMKNNYRNAHSAATTWSGQY 120
DB 64 gAltgtYesaVgnaesryvltgrYdsapAtdgsctAlgwtVmkNnyrNahsaAttwsGqy 123
QY 121 VGGAERINTOMLITSGTEBANAMKSTLVGHDTFTVKRPSAASIDAKKAGVNGNPIDA 180
DB 124 vGgaerIntomLiTsGteBanamKstlvGHdtftVkrPsAaSIdaKkAgvNgnpIda 183
QY 181 VQG 183
DB 181 vqg 186
RESULT 9
Y80513
ID Y80513 standard; Protein: 184 AA.
XX
XX Y80513;
XX
XX 06-JUN-2000 (first entry)
XX
XX Streptomyces avidinII mat protein.
XX
XX Plant somatic tissue degeneration; plant essential factor; depletion;
XX viability; mat gene; plant development; plant morphology; flower;
XX fruit plant.
XX
XX Streptomyces avidinII.
XX
XX WO200007427-A2.
XX
XX 17-FEB-2000.
XX
```

XX 30-JUL-1999; 99WO-IL00420.
PF
XX 03-AUG-1998; 98IL-0125632.
PR
XX (AGRI-) AGRIC RES ORG.
PA
XX Kapulnik Y, Ginzberg I;
PI
XX WPI: 2000-195402/17.
DR N-PSDB; 291074.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
PS
XX Examples; Page 85; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the streptomycetes
CC avidin11 mst protein as an example of a heterologous protein introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 184 AA;

Query Match 99.5%; Score 182; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 7.4e-172;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRIIVAAIAVSTITVSTRASADSPSKDSKAOVSAEAGITGTWNLQGSTITVAGADG 61
DB 3 KRIIVAAIAVSTITVSTRASADSPSKDSKAOVSAEAGITGTWNLQGSTITVAGADG 62
OY 62 ALTGYESAVGNAESRYVLTGRYDSAPATDGSFTALGWTVAKNMYRNAHSATTWSGOTV 121
DB 63 altgyesavvnaesryvltgrydsapatdgsftalgtvawknyrnhsattwsqyv 122
OY 122 GGAEARINTQWLLTSGTTEANAMKSTLVGHDPFTFKVPSAASIDAAKKAGVNNGNPLDAV 181
DB 123 ggaearintqwltsgteanaawkstlvghdtftkvkpsaasidaakkagvnngnpldav 182
OY 182 QQ 183
DB 183 qq 184

RESULT 10
R50515
ID Y80515 standard; Protein: 161 AA.
XX
XX Y80515;
AC
XX
XX 06-JUN-2000 (first entry)
DT
XX Streptomyces avidin11 mprost protein.
DE
XX Streptomyces avidin11 mprost protein.
DE
XX Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; mprost gene; plant development; plant morphology; flower;
KW fruit plant.
XX
XX Streptomyces avidin11.
OS
XX
XX WO200007427-A2.
PN

XX 17-FEB-2000.
PD
XX 30-JUL-1999; 99WO-IL00420.
PF
XX 03-AUG-1998; 98IL-0125632.
PR
XX (AGRI-) AGRIC RES ORG.
PA
XX Kapulnik Y, Ginzberg I;
PI
XX WPI: 2000-195402/17.
DR N-PSDB; 291076.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
PS
XX Examples; Page 87; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the streptomycetes
CC avidin11 mprost protein as an example of a heterologous protein
CC introduced into the plants. The methods can provide for the selective
CC and optionally reversible cell degeneration in somatic plant tissue.
CC They can be used for artificially controlling plant development and
CC morphology. They can be used e.g. to decrease the number of flowers
CC in fruit producing plants so as to increase the number of fruits which
CC reach maturity.
XX
SQ Sequence 161 AA;

Query Match 87.4%; Score 160; DB 21; Length 161;
Best Local Similarity 100.0%; Pred. No. 3.2e-150;
Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 ADPSKDSKAQVSAEAGITGTWYNLQGSTFTVAGADGALTGYESAVGNAESRYVLTGR 83
DB 2 adpskdskaqvsaeeagltgtwnqgstftvtagadgaltgyesavvnaesryvltgr 61
OY 84 YDSAPATDGSFTALGWTVAKNMYRNAHSATTWSGQYVGGAERINTQWLLTSGTTEANA 143
DB 62 ydsapatdgsftalgtvawknyrnhsattwsqyvggaearintqwltsgteana 121
OY 144 WKSTLVGHDPFTFKVPSAASIDAAKKAGVNNGNPLDAVQQ 183
DB 122 wkstlvghdtftkvkpsaasidaakkagvnngnpldavqq 161

RESULT 11
R56483
ID R56483 standard; Protein: 435 AA.
XX
XX R56483;
AC
XX
XX 26-MAR-1995 (first entry)
DT
XX SCFV PRAS108 and PRAS112.
DE
XX Amplification; single chain variable region fusion protein; PCR.
KW
XX
XX Synthetic.
OS
XX
XX WO9415644-A.
PN
XX
XX 21-JUL-1994.
PD
XX

CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.

XX Sequence 162 AA;

Query Match 86.9%; Score 159; DB 21; Length 162;

Best Local Similarity 100.0%; Pred. No. 3.1e-149; Mismatches 0; Indels 0; Gaps 0;

Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 DPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALGTGESAVGNAESRYVLTGRY 84
DB 4 dpskdskaqysaaegitgtwnglqstflvtagadgaltgyesavgnaesryvltgry 63
OY 85 DSAPATDGGSTALGWTVAMKNNYRNAHSATWMSGQYVGGAEARINTQWLITSGTTEANAW 144
DB 64 dsapatdgsqstalgwvawknynrhaasatwsgqyvggaearintqwlitsgtteanaw 123
OY 145 KSTLVGHDFTRKPSAASIDAKKAGVNNGNPLDAVQ 183
DB 124 kstlvghdftkvpasasidaakkagvnngnpldavq 162

RESULT 14

P80160
ID P80160 standard; protein; 163 AA.

AC P80160;

DT 13-OCT-1990 (first entry)

DE Biosynthetic protein with streptavidin trailer.

XX Biosynthetic Antibody Binding site (BABS): variable domain;
KM anti-CEA monoclonal antibody; streptavidin.

XX Synthetic.

XX W08809344-A.

PN 01-DEC-1988.

PD 19-MAY-1988; 88WO-US01737.

PF 21-MAY-1987; 87US-0052800.

PR (CREA-) CREATIVE BIOMOLECULES INC.

PI Huston JS, Oppermann H;

DR WPI; 1988-353928/49.

DR N-PSDB; N80186.

XX Recombinant multifunctional protein -
XX having an antibody binding site and a sequence for biological activity,
XX ion sequestering or binding to a solid support.

XX Disclosure; 115pp; English.

XX Multi-functional biosynthetic protein comprising single
CC chain BABS and streptavidin protein trailer linked via a spacer

CC See also N80171-N80192.

XX Sequence 163 AA;

Query Match 86.9%; Score 159; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.2e-149;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 DPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALGTGESAVGNAESRYVLTGRY 84
DB 5 dpskdskaqysaaegitgtwnglqstflvtagadgaltgyesavgnaesryvltgry 64
OY 85 DSAPATDGGSTALGWTVAMKNNYRNAHSATWMSGQYVGGAEARINTQWLITSGTTEANAW 144
DB 65 dsapatdgsqstalgwvawknynrhaasatwsgqyvggaearintqwlitsgtteanaw 124
OY 145 KSTLVGHDFTRKPSAASIDAKKAGVNNGNPLDAVQ 183
DB 125 kstlvghdftkvpasasidaakkagvnngnpldavq 163

RESULT 15

ID P70492 standard; protein; 182 AA.

AC P70492;

DT 06-MAR-1991 (first entry)

DE Streptavidin sequence.

XX Streptavidin; N-terminal; fusion gene; fusion protein;

XX Streptomyces avidinii.

PN W08705026-A.

PD 27-AUG-1987.

PF 24-FEB-1987; 87WO-US00397.

PR 24-FEB-1986; 86US-0833324.

PA (UYCO-) COLUMBIA UNIV N Y.

PA (UYNV-) UNIV NEW YORK.

PI Cantor CR, Axel R, Garana C;

DR WPI; 1987-250198/35.

DR N-PSDB; N70810.

PT DNA encoding streptavidin - obtd. by restriction endo-nuclease
XX digestion of chromosomal DNA of Streptomyces avidinii

XX Disclosure; Figure 3; 54pp; English.

XX The sequence is that of streptavidin from *S. avidinii*. Streptavidin
CC may be expressed from a fusion gene comprising its coding gene and DNA
CC encoding a target protein of interest, where the streptavidin has
CC binding sites for biotin or deriv. The streptavidin has 4 binding
CC sites free for biotin, and is produced free of biotin
CC contamination. Improved streptavidins may also be produced by
CC site-directed mutagenesis. The fused gene may be used to produce
CC labelled, chemically-modified proteins in vivo, and to isolate
CC proteins when only the sequence of the gene is known.

XX Sequence 182 AA;

Query Match 83.1%; Score 152; DB 8; Length 182;
Best Local Similarity 100.0%; Pred. No. 2.8e-142; Mismatches 0; Indels 0; Gaps 0;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 AQVSAEAGITGTWYNQLGSTFIYTAGADGALGTGESAVGNAESRYVLTGRYSAPARD 91
DB 31 aqvsaeeagltgtwnglqstflvtagadgaltgyesavgnaesryvltgrydsapard 90
OY 92 GSGTALGWTVAMKNNYRNAHSATWMSGQYVGGAEARINTQWLITSGTTEANAWKSTLVGH 151
DB 91 gsgtalgwvawknynrhaasatwsgqyvggaearintqwlitsgtteanawkstlvgh 150

Fri Apr 6 08:40:59 2001

us-09-589-870-2.rag

Page 9

QY 152 DTFTKVPSPAASIDAKKAGVNNGNPLDAVQ 183
|||
Db 151 dftfkvpspaasidaakkagvnnngnpldavq 182

Search completed: April 6, 2001, 01:25:30
Job time: 2004 sec

THIS PAGE BLANK (USPTO)

67 rgluserAlaValGlyValAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||
250 CGAGTGGCGCGTGGCCACGCCGACGAGCCCTACGTCCTACCGGTCGT 299
84 YrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPThr 100
|||||
300 ACGACAGCGCGCCGCGCACCGGACGCGGACCGCCCTCGGTGGAGC 349
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTpsE 117
|||||
350 GTGGCTCGGAAGATACTACCCGACGCCCTCGCCGACGACGAGGAG 399
117 rglYglNtYrValGlyGlyValaGluAlaArgIleAsnThrGlnTyrLeuL 134
|||||
400 CGGCGACGACGTCGGCGCGCCGACGCGAGGATCAACACCCAGTGGCTGC 449
134 eutHrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCCGGCACCGACGAGCCGCTGGAAGTCCACGCTGGTCGGC 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaI 167
|||||
500 CACGACACCTTCACCAAGGTAAGCGCTCGCGCTCCATCGACGCGGC 549
167 alySLysAlaGlyValaLysAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 GAAGAAAGCGCGCGCTCAACACGCAACCCGCTCGACGCCGTTCAAGAG 598

seq_name: gb_pat1:A93649

seq_documentation_block:

LOCUS A93649 638 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0799890.ACCESSION A93649
VERSION A93649.1 GI:6741838KEYWORDS
SOURCE unidentified.
ORGANISM unidentifiedREFERENCE
1 (bases 1 to 638)

AUTHORS Mueller, R.D. and Deger, A.D.

TITLE Recombinant inactive core streptavidin mutants

JOURNAL Patent: EP 0799890-A 08-OCT-1997;
BOEHRINGER MANNHEIM GMBH (DE)FEATURES
Location/Qualifiers
1..638sig_peptide /organism="unidentified"
/db_xref="taxon:32644"CDS 50..121
50..601
/note="unnamed protein product"mat_peptide /codon_start=1
/protein_id="CAB69587.1"

BASE COUNT 115 a 244 c 193 g 86 t

ORIGIN 122..598
AASIDAAKAGVNGNPLDAVQO*alignment_scores:
Quality: 183.00 Length: 183
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000alignment_block:
US-09-589-870-2 x A93649 ..

Align seg 1/1 to: A93649 from: 1 to: 638

1 MetArgLysIleValValaIleAlaValSerLeuThrValSe 17
|||||
50 ATGCGCAGATCGTCGTCGACCATGCGCTTCCCTGACACAGGCTGC 99
17 rleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGln 34
|||||
100 GATTACGGCCAGCGCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAGG 149
34 aLserAlaIaGluAlaGlyIleThrGlyThrTyrPThrAsnGlnLeuGly 50
|||||
150 TCTCGCGCGCGGAGCGCGGATCACCGCACCTGGTAAACACGAGCTCGC 199
51 SerThrPheIleValThrAlaGlyAlaAspGlyValaLeuThrGlyThrTyr 67
|||||
200 TCGACCTTCATCGTACCGCGCGGCGCGGAGGAGGAGGAGGAGGAGGAG 249
67 rgluserAlaValGlyValaLysAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||
250 CGAGTGGCGCGTGGCCACGCCGAGAGCCGCTACGTCCTACCGGTCGT 299
84 YrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPThr 100
|||||
300 ACGACACGCGCCCGCGCACCGGACGCGGACCGCCCTCGGTGGAGC 349
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTpsE 117
|||||
350 GTGGCTCGGAAGATACTACCCGACGCCCTCGCGCTCCATCGACGCGGC 399
117 rglYglNtYrValGlyGlyValaGluAlaArgIleAsnThrGlnTyrLeuL 134
|||||
400 CGGCGACGACGTCGGCGCGCCGACGCGGAGGATCAACACCCAGTGGCTGC 449
134 eutHrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCCGGCACCGACCGAGGCGCACGCTGGAAGTCCACGCTGGTCGGC 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaI 167
|||||
500 CACGACACCTTCACCAAGGTAAGCGCTCGCGCTCCATCGACGCGGC 549
167 alySLysAlaGlyValaLysAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 GAAGAAAGCGCGCGTCAACACGCAACCCGCTCGACGCCGTTCAAGAG 598

seq_name: gb_pat1:A39565

seq_documentation_block:

LOCUS A39565 1356 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 4 from Patent WO9415644.ACCESSION A39565
VERSION A39565.1 GI:2295847KEYWORDS
SOURCE unidentified.
ORGANISM unidentifiedREFERENCE
1 (bases 1 to 1356)

AUTHORS Epenetos, A.A., Spooner, R.A. and Deonarain, M.

TITLE COMPOUNDS FOR TARGETING

JOURNAL Patent: WO 9415644-A 4 21-JUL-1994;
IMP CANCER RES TECH (GB)

COMMENT Other publication GB 2289679 951129.

FEATURES
Location/Qualifiers
1..1356sig_peptide /organism="unidentified"
/db_xref="taxon:32644"CDS 40..1347
/note="unnamed protein product"mat_peptide /codon_start=1
/protein_id="CAA02449.1"BASE COUNT 400 a 344 c 229 g 183 t
ORIGIN 1
KASGYTPTSYMHVWKQRPGRGLMIGRIDPNSGTYKNEFKSKATITVDKPSSTAV

MOLESITSEDSAVVYICARYDYSSYFDYWGQTTVTYSSGGGSGGGGSGGGSQAV
VTOESALTPSGEIVTLTGRSSTGAVTTSNVMWQERPDHLFTGLIGETNNRAGVP
ARFSGSLIGDKAALITGAQOTDEALFYFCALVSNHWMWGGLTGLVLEAPAAPA
DPSKDSKAOVSAEAEGITGTWYNOLGSPFIYTAGADALGTYESVGNABERYVLTG
RTDSAPATDGSSTALGWTAVAMNNYRNMAHSATTTWGGQYVGAEBARINTQWLTSCTTE
ANAMKSTLVGHDTFTKVRKPSAASIDAARKAGVNNGNPLDAVQO"

BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN

alignment_scores:
Quality: 160.00 Length: 160
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x A39565 ..

Align seg 1/1 to: A39565 from: 1 to: 1356

24 AAlaasproSerLysaspSerLysAlaGlnValSerAlaAlaGlnValAlaGln 40
|||||
865 GCAGACCCGCTCCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGGAAGCTGG 914
40 YLleThGlyThrTrpTyraSngInLeuGlySerThrPheIleValThra 57
|||||
915 TATACCTGGACCTGGTGTATACCAACTGGGCTGCACCTTCTTTGTGACCG 964
57 laGlyAlaAspGlyAlaLeuThrGlyThrTyrgLuserAlaValAlaGln 73
|||||
965 CTGGTGGGAGAGGAGCTGTGACCTGCACATCGATCTGCGGTTGGTGAAC 1014
74 AAlaGlnSerArgTyraValLeuThrGlyArgTyraSpsSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACGGCCGTTATGACTCTGCACCTGCCAC 1064
90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsn 107
|||||
1065 CGATGGCTCTGTACCGCTCTGGGCTGGACTGTGGCTTTGGAAAAACAAC 1114
107 yTrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyraValGly 123
|||||
1115 ATCGTAATGCCACACAGGCCACGTACGTGGCTGGCCAAATACGTTGGCGGT 1164
124 AAlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGln 140
|||||
1165 GCTGAGGCTGCTATCAACACTCAGTGGCTGTTAACTCCGGCAGCTACCGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlnHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAG 1264
157 aLysProSerAlaAlaSerIleAspAlaAlaLysAlaGlnValAsn 173
|||||
1265 TTAAAGCCTTGTGCTAGCATGATGCTGCCAAGAAAGAGGGGTAAC 1314
174 AsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
1315 AACGTTAACCTCTTAGACGGCTTTCAGCAA 1344

seq_name: gb_pat1:A93152

seq_documentation_block:

LOCUS A93152 1356 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 4 from Patent EP0815872.
ACCESSION A93152
VERSION A93152.1 GI:6741540
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos, A.A. and Deonaraia, M.

TITLE Compounds for targeting
JOURNAL Patent: EP 0815872-A 07-JAN-1998;
IMP CANCER RES TECH (GB)
FEATURES
source Location/Qualifiers
1..1356
/organism="unidentified"
/db_xref="taxon:32644"
40..1347
/note="unnamed protein product"

CDS

/codon_start=1
/protein_id="CAB69540.1"
/db_xref="GI:6741541"
/translation="MKYLPTRAAGLLLAQAPAMAQVQLDQAPAEIVPGASVKLSC
KASGYPTFSWMHWKORPGRLEWIGRIDPNSGGTKYERKSRATLTVDRPSSTAY
VTOESALTPSGEIVTLTGRSSTGAVTTSNVMWQERPDHLFTGLIGETNNRAGVP
ARFSGSLIGDKAALITGAQOTDEALFYFCALVSNHWMWGGLTGLVLEAPAAPA
DPSKDSKAOVSAEAEGITGTWYNOLGSPFIYTAGADALGTYESVGNABERYVLTG
RTDSAPATDGSSTALGWTAVAMNNYRNMAHSATTTWGGQYVGAEBARINTQWLTSCTTE
ANAMKSTLVGHDTFTKVRKPSAASIDAARKAGVNNGNPLDAVQO"

BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN

alignment_scores:
Quality: 160.00 Length: 160
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x A93152 ..

Align seg 1/1 to: A93152 from: 1 to: 1356

24 AAlaasproSerLysaspSerLysAlaGlnValSerAlaAlaGlnValAlaGln 40
|||||
865 GCAGACCCGCTCCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGGAAGCTGG 914
40 YLleThGlyThrTrpTyraSngInLeuGlySerThrPheIleValThra 57
|||||
915 TATACCTGGACCTGGTGTATACCAACTGGGCTGCACCTTCTTTGTGACCG 964
57 laGlyAlaAspGlyAlaLeuThrGlyThrTyrgLuserAlaValAlaGln 73
|||||
965 CTGGTGGGAGAGGAGCTGTGACCTGCACCTACGAATCTGCCGTTGGTGAAC 1014
74 AAlaGlnSerArgTyraValLeuThrGlyArgTyraSpsSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACGGCCGTTATGACTCTGCACCTGCCAC 1064
90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsn 107
|||||
1065 CGATGGCTCTGTACCGCTCTGGGCTGGACTGTGGCTTTGGAAAAACAAC 1114
107 yTrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyraValGly 123
|||||
1115 ATCGTAATGCCACACGCCCTACGTGGCTGGCCAAATACGTTGGCGGT 1164
124 AAlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGln 140
|||||
1165 GCTGAGGCTGCTATCAACACTCAGTGGCTGTTAACTCCGGCAGCTACCGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlnHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAG 1264
157 aLysProSerAlaAlaSerIleAspAlaAlaLysAlaGlnValAsn 173
|||||
1265 TTAAAGCCTTGTGCTAGCATGATGCTGCCAAGAAAGAGGGGTAAC 1314
174 AsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
1315 AACGTTAACCTCTTAGACGGCTTTCAGCAA 1344

seq_name: gb_pat1:AR082490

seq_documentation_block:

LOCUS AR082490 1356 bp DNA PAT 31-AUG-2000

DEFINITION Sequence 4 from patent US 5973116.

ACCESSION AR082490

VERSION AR082490.1 GI:10009216

KEYWORDS

SOURCE

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1356)

AUTHORS Epetetos,A.,Antonlou, Spooner,R.,Anthony and Deonarin,M.

TITLE Compounds for targeting

JOURNAL Patent: US 5973116-A 4 26-OCT-1999;

FEATURES

source 1..1356

BASE COUNT 332 a 357 c 361 g 306 t

ORIGIN

alignment_scores:

Quality: 160.00 Length: 160

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x AR082490 ..

Align seg 1/1 to: AR082490 from: 1 to: 1356

24 AlaasproserlysaspserysAlaGlnValSerAlaAlaGlnAlaG1 40

|||||

865 GCAGACCCGTCAGGAGCTCCAAAGCTCAGGTTCTGCACCCGAACTGG 914

40 yilethg1yThrTyrAsnGlnLeuGlySerThrPheIleValThra 57

|||||

915 TATCAGCTGGACCTGCTATTAACCACTGGGGTCACCTTCATTGTGACCG 964

57 laGlyAlaaspg1yAlaLeuThrg1yThrTyrGlnSerAlaValGlyAsn 73

|||||

965 CTGGTGGGAGCGAGCTCTGACTGGCCTTATGACTCTGCCTGCTGTAAC 1014

74 AlaGlnSerArgTyrValLeuThrg1yArgTyrAspSerAlaProAlaTh 90

|||||

1015 GCAGAAATCCGCTACGTACTGAGCGCTTATGACTCTGCACCTGCCAC 1064

90 rAspGlySerG1yThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107

|||||

1065 CGATGGCTCTGGTACCCGCTCTGGGCTGGACTGTGGCTTGGAAAAACAAC 1114

107 yTrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGly 123

|||||

1115 ATCGTATGGGACACAGCGCACCTACGTGCTGGCCAAATGAGTGGCGGT 1164

124 AlaGlnAlaArgIleAsnThrg1nTrpLeuLeuThrSerGlyThrThrg1 140

|||||

1165 GCTGAGGCTGTATACACTCACTGAGTGGCTTTAAACATCCGCGCATACGA 1214

140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157

|||||

1215 AGCGAAATGCATGGAAATCGACACTAGTAGTCATGACACCTTTTCCAAAG 1264

157 aLysProSerAlaAlaSerIleAspAlaAlaLysIlySAlaGlyValAsn 173

|||||

1265 TTAAAGCTTCTGCTGCTAGCATGTATGCTGCAAGAAAGCAGCGGTAAAC 1314

174 AsnGlyAsnProLeuAspAlaValGlnGln 183

|||||

1315 AACGTTAACCTCTAGACGCTGTTCAGCAA 1344

seq_name: gb_pat1:A93838

seq_documentation_block:

LOCUS A93838 498 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 7 from Patent WO9728263.

ACCESSION A93838

VERSION A93838.1 GI:6741969

KEYWORDS

SOURCE

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 498)

AUTHORS Lubitz,W. and Sleytr,U.

TITLE RECOMBINANT EXPRESSION OF S-LAYER PROTEINS

JOURNAL Patent: WO 9728263-A 07-AUG-1997;

FEATURES

source 1..498

BASE COUNT 120 a 137 c 131 g 110 t

ORIGIN

alignment_scores:

Quality: 159.00 Length: 159

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x A93838 ..

Align seg 1/1 to: A93838 from: 1 to: 498

25 AspProserlysaspserysAlaGlnValSerAlaAlaGlnAlaG1 41

|||||

7 GACCCGTCCAAAGAGCTCCAAAGCTCAGGTTCTGCACCCGAACTGGTAT 56

41 eThrG1yThrTyrAsnGlnLeuGlySerThrPheIleValThraLag 58

|||||

57 CACTGGCACTGCTATTAACCACTGGGGTGCACCTTCTATTGTGACCCCTG 106

58 lYalAaspg1yAlaLeuThrg1yThrTyrGlnSerAlaValGlyAsnAla 74

|||||

107 GTGGGAGAGGAGCTCTGACTGGCCTACGAAATCTGCGTTGGTAAACGA 156

75 GlnSerArgTyrValLeuThrg1yArgTyrAspSerAlaProAlaThras 91

|||||

157 GAATCCCGCTACGTACTGAGCGCTTATGACTCTGCACCTGCCACGA 206

91 pGlySerG1yThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyrA 108

|||||

207 TGCTCTGGTACCCGCTCTGGGCTGGACTGTGGCTTGGAAAAACAACATATC 256

108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyAla 124

|||||

257 GTATGGGACACAGCGCACCTACGTGCTGGCCAAATGAGTGGCGGTGCT 306

125 GlnAlaArgIleAsnThrg1nTrpLeuLeuThrSerGlyThrThrg1ua1 141

|||||

307 GAGCTGTGTATCAACACTCACTGAGTGGCTTTAAACATCCGCGCATACGAGC 356

141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158

|||||

357 GAATGCATGGAAATCGACACTAGTAGTCATGACACCTTTTACAAATTA 406

158 ySProSerAlaAlaSerIleAspAlaAlaLysIlySAlaGlyValAsnAsn 174

|||||

407 AGCTTCTGCTGCTAGCATGTATGCTGCCAAGAAAGCAGCGGTAAAC 456

175 GlyAsnProLeuAspAlaValGlnGln 183

|||||

457 GTTAACCTCTAGACGCTGTTCAGCAA 483

```
seq_name: gb_pat1:109059
seq_documentation_block:
LOCUS       109059             498 bp                PAT
DEFINITION  Sequence 38 from Patent WO 8809344.
ACCESSION   109059
VERSION     109059.1  GI:588242
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 498)
AUTHORS     Huston J.S. and Oppermann H.
TITLE       TARGETED MULTIFUNCTIONAL PROTEINS
JOURNAL     Patent: WO 8809344-A 38 01-DEC-1988;
FEATURES
source      1..498
            /organism="unknown"
BASE COUNT  117 a 137 c 133 g 111 t
ORIGIN
alignment_scores:
Quality: 159.00      Length: 159
Ratio: 1.000         Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-09-589-870-2 x 109059  ..
Align seg 1/1 to: 109059 from: 1 to: 498
25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI1 41
13 GACCCGTCGAAGGACTCCAAAGCTCAGGTTTCTGCTGCCGAAGCTGGTAT 62
41 eThhGlyThrTyrTrpTyrAsnGlnLeuGlySerThrPheIleValThrIaG 58
63 CACTGGCAGCCTGGTATACCAACTGGGGTGCACCTTTCATTGTGTGACCCCTG 112
58 LysAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValAlaGlyAsnAla 74
113 GTGGGAGCGAGCTCTGACTGCGACCTACGAATCTGGCGTTGGTGAACGCA 162
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
163 GAATCCCGCTACGACTGACTGGCGCTTATGACTCTGCACCTGCCACGCA 212
91 pGlySerGlyThrAlaLeuGlyTTPthValAlaTTPlysAsnAsnTyrA 108
213 TGGCTCTGGTACCGCTCTGGCTGGACTGTGCTTGGAAAAACAATATC 262
108 rGaAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValAlaGlyAla 124
263 GTAATGGCGACAGCGCCACTACGTGCTGGCCAAATACGTTGGCGGTCT 312
125 GluAlaArgIleAsnThrGlnTTPLeuLeuThrSerGlyThrThrGluAl 141
313 GAGCGCTGTATCAACACTGACTGCTGTAAATCCGCGACACTACCGAAGC 362
141 eAsnAlaLeuTTPlysSerThrLeuValAlaGlyHisAspThrPheThrIysValL 158
363 GAATGCAATGGAAATCGACACTAGTAGTCATGACACCTTTACCAAGTTA 412
158 ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
413 AGCCTTCTGCTGCTAGCATGTGATCTGCGCAAGAAAGCAGCGGTAAACAC 462
175 GlyAsnProLeuAspAlaValAlaGlnI1n 183
463 GGTAACCTCTAGACGCTGTTCAGCAA 489
```

```
seq_name: gb_pat1:A00743
seq_documentation_block:
LOCUS       A00743             507 bp                DNA                PAT
DEFINITION  S.avidinii synthetic gene for streptavidin.
ACCESSION   A00743
VERSION     A00743.1  GI:14605
KEYWORDS
SOURCE      Streptomyces avidinii.
ORGANISM    Streptomyces avidinii.
REFERENCE   1 (bases 1 to 507)
AUTHORS     Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
TITLE       Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
JOURNAL     Edwards R.
SYNTHETIC GENE
PATENT: WO 8903422-A 2 20-APR-1989.
JOURNAL     Bio-Technology Ltd
FEATURES
source      1..507
            /organism="Streptomyces avidinii"
            /db_xref="taxon:1895"
            /note="synthetic gene"
            /codon_start=1
            /transl_table=11
            /product="streptavidin"
            /protein_id="CAA0084.1"
            /db_xref="GI:14606"
            /translation="MDPSKSKAOYSAAEAGITGMYNOLGSTFTVTAGADALNTGY
            ESAVGNASERYVLVTRYSAPATDGSGLMGVTVMKNRYRNAHSATTWGQYVGAE
            ARINQWLLTSGTTEANAMKSTLVGHDFPTVKPSAASIDAKKAGVNNGNPLDAVQO"
BASE COUNT  125 a 137 c 130 g 115 t
ORIGIN
alignment_scores:
Quality: 159.00      Length: 159
Ratio: 1.000         Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-09-589-870-2 x A00743  ..
Align seg 1/1 to: A00743 from: 1 to: 507
25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI1 41
13 GACCCGTCGAAGGACTCCAAAGCTCAGGTTTCTGCTGCCGAAGCTGGTAT 62
41 eThhGlyThrTyrTrpTyrAsnGlnLeuGlySerThrPheIleValThrIaG 58
63 CACTGGCAGCCTGGTATACCAACTGGGGTGCACCTTTCATTGTGTGACCCCTG 112
58 LysAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValAlaGlyAsnAla 74
113 GTGGGAGCGAGCTCTGACTGCGACCTACGAATCTGGCGTTGGTGAACGCA 162
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
163 GAATCCCGCTACGACTGACTGGCGCTTATGACTCTGCACCTGCCACGCA 212
91 pGlySerGlyThrAlaLeuGlyTTPthValAlaTTPlysAsnAsnTyrA 108
213 TGGCTCTGGTACCGCTCTGGCTGGACTGTGCTTGGAAAAACAATATC 262
108 rGaAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValAlaGlyAla 124
263 GTAATGGCGACAGCGCCACTACGTGCTGGCCAAATACGTTGGCGGTCT 312
125 GluAlaArgIleAsnThrGlnTTPLeuLeuThrSerGlyThrThrGluAl 141
313 GAGCGCTGTATCAACACTGACTGCTGTAAATCCGCGACACTACCGAAGC 362
```

141 aasnaIatrpIySerThrLeuValGlyHisAspThrPheThrIysValL 158
|||||
363 GAATGCATGGAATGCACACTAGTATGATGACCTTACCAAGTTA 412
158 yspProSerAlaIaSerIleAspAlaIaIysAlaGlyValaAsn 174
|||||
413 AGCCTTCTGCTGCTAGCATTTGATGTCACCAAGAAAGAGCGTAAACAC 462
175 GLyAsnProLeuAspAlaValGlnGln 183
|||||
463 GGTAAACCTCTAGACGCTGTTCAGCAA 489

seq_name: gb_pat1:A00744

seq_documentation_block:

LOCUS A00744 507 bp DNA PAT 28-JAN-1993
DEFINITION S.avidinii synthetic gene (reverse complement) for streptavidin.

ACCESSION A00744
VERSION A00744.1 GI:14607

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

BASE COUNT

ORIGIN

alignment_scores:

Quality: 159.00 Length: 159

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x A00744/rev ..

Align seg 1/1 to reverse of: A00744 from: 1 to: 507

25 AspProSerIysAspSerIysAlaGlnValSerAlaIaGlyI 41
|||||
495 GACCGGTCCAAAGACTCCAAAGCTAGTTCTGCGAGCGCAAGCGGTAT 446
41 eThrGlyThrTyrTrpTyrAsnGlnLeuGlySerThrPheIleValThrIaG 58
|||||
445 CACTGGCACTGCTGATTAACCAACTGGGCTGCACTTTCATTTGTGACCGCTG 396
58 IyAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAla 74
|||||
395 GTGGGAGGAGGAGCTGACTGGCAGCTGAGATCTGCGGTTGGTAAACCA 346
75 GluSerArgTyrValIleuThrGlyArgTyrAspSerAlaProAlaThrAs 91
|||||
345 GAATCCCGCTAGTACTGACTGCGGTTATGACTGTCACCTGCCACCGCA 296
91 pGlySerGlyThrAlaLeuGlyTyrThrValAlaIaIaIaIaIaIaIaIaIa 108
|||||
295 TGGCTCTGTGATCCGCTGCGGCTGAGCTGTGGTGAAGAAACAACTATC 246
108 rGAsnAlaHisSerAlaThrThrTyrSerGlyGlnTyrValGlyIyAla 124
|||||
245 GTPAATGGCAGACGCGCACTAGCTGCTGCGCAATACGTTGGCGGTCT 196

125 GluAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrGluAl 141
|||||
195 GAGGCTCGATACACTCAGTGGCTGTTAATACCGGCACTACCGAAGC 146
141 aasnaIatrpIySerThrLeuValGlyHisAspThrPheThrIysValL 158
|||||
145 GAATGCATGGAATGCACACTAGTATGATGACCTTACCAAGTTA 96
158 yspProSerAlaIaSerIleAspAlaIaIysAlaGlyValaAsn 174
|||||
95 AGCCTTCTGCTGCTAGCATTTGATGTCACCAAGAAAGAGCGTAAACAC 46
175 GLyAsnProLeuAspAlaValGlnGln 183
|||||
45 GGTAAACCTCTAGACGCTGTTCAGCAA 19

seq_name: gb_pat1:A20698

seq_documentation_block:

LOCUS A20698 525 bp DNA PAT 22-AUG-1994

DEFINITION Fxa-StrpA DNA sequence.

ACCESSION A20698

VERSION A20698.1 GI:583400

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

BASE COUNT

ORIGIN

alignment_scores:

Quality: 159.00 Length: 159

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x A20698 ..

Align seg 1/1 to: A20698 from: 1 to: 525

25 AspProSerIysAspSerIysAlaGlnValSerAlaIaGlyI 41
|||||
37 GACCGGTCCAAAGACTCCAAAGCTAGTTCTGCGAGCGCAAGCGGTAT 86
41 eThrGlyThrTyrTrpTyrAsnGlnLeuGlySerThrPheIleValThrIaG 58
|||||
87 CACTGGCACTGCTGATTAACCAACTGGGCTGCACTTTCATTTGTGACCGCTG 136
58 IyAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAla 74
|||||
137 GTGGGAGGAGGAGCTGACTGGCAGCTAGCAATCTGCGGTTGGTAAACCA 186
75 GluSerArgTyrValIleuThrGlyArgTyrAspSerAlaProAlaThrAs 91
|||||
187 GAATCCCGCTAGTACTGACTGCGGTTATGACTGTCACCTGCCACCGCA 236
91 pGlySerGlyThrAlaLeuGlyTyrThrValAlaIaIaIaIaIaIaIaIaIa 108
|||||
237 TGGCTCTGTGATCCGCTGCGGCTGAGCTGTGGTGAAGAAACAACTATC 286
108 rGAsnAlaHisSerAlaThrThrTyrSerGlyGlnTyrValGlyIyAla 124
|||||
287 GTPAATGGCAGACGCGCACTAGCTGCTGCGCAATACGTTGGCGGTCT 336

```

125 GUAlAAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
|||||
337 GAGGCTGTATCAACACTCAAGTGGCTTTAACTCCGGCACTACCGAAGC 386
|||||
141 AAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158
|||||
387 GAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAACTTA 436
|||||
158 YSPROSERAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
437 AGCCTTCTGCTGCTAGCATGATGCTGCCAAGAAACGAGCGGTAAACAC 486
|||||
175 GLYAsnProLeuAspAlaValGlnGln 183
|||||
487 GGTAACTCTTAGAGCTGTTCAGCAA 513
|||||

seq_name: gb_pat1:115647

seq_documentation_block:
LOCUS R15647 525 bp DNA PAT 02-APR-1996
DEFINITION Sequence 2 from patent US 5470573.
ACCESSION I15647
VERSION I15647.1 GI:1250555
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 525)
AUTHORS Lubitz,W. and Szostak,M.P.
TITLE Immunoens comprising the non-lytic membrane spanning domain of
JOURNAL Patent: US 5470573-A 2 28-NOV-1995;
FEATURES
source 1..525
location/Qualifiers
BASE COUNT 128 a 140 c 138 g 119 t
ORIGIN

alignment_scores:
Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x I15647 ..

Align seg 1/1 to: I15647 from: 1 to: 525

```

```

125 GUAlAAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
|||||
337 GAGGCTGTATCAACACTCAAGTGGCTTTAACTCCGGCACTACCGAAGC 386
|||||
141 AAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158
|||||
387 GAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAACTTA 436
|||||
158 YSPROSERAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
437 AGCCTTCTGCTGCTAGCATGATGCTGCCAAGAAACGAGCGGTAAACAC 486
|||||
175 GLYAsnProLeuAspAlaValGlnGln 183
|||||
487 GGTAACTCTTAGAGCTGTTCAGCAA 513
|||||

seq_name: gb_pat1:101349

seq_documentation_block:
LOCUS I01349 638 bp ss-DNA PAT 21-MAY-1993
DEFINITION Sequence 2 from Patent US 4839293.
ACCESSION I01349
VERSION I01349.1 GI:270135
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 638)
AUTHORS Cantor,C.R., Axel,R. and Argarana,C.
TITLE DNA encoding streptavidin, streptavidin produced therefrom, fused
polypeptides which include amino acid sequences present in
JOURNAL Patent: US 4839293-A 2 13-JUN-1989;
The Trustees of Columbia University in the City of New York;
FEATURES
source 1..638
location/Qualifiers
BASE COUNT 114 a 244 c 194 g 86 t
ORIGIN

alignment_scores:
Quality: 152.00 Length: 152
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x I01349 ..

Align seg 1/1 to: I01349 from: 1 to: 638

```

```

115 htrtperglyntyrvalglyalaaglualaatrtglestnrgin 131
|||||
393 CGTGGAGCGCCAGTACGTGGCGGGCCGAGCCAGATCAACACCAG 442
|||||
132 tptlenenthrserglythrthrtrigualaasnaatrtplysertthle 148
|||||
443 TGCGTGTGTGACCTCCGGCACACCGAGGCCACGCTGTGAATCTCACGCT 492
|||||
148 uvalglyhisapthrphenrhrlysalysproserataalaserilea 165
|||||
493 GGTGGCCACGACACTTCACAAAGGTGAAGCCGTCCGGCGCTCATCG 542
|||||
165 spataalalysalaglyvalalasnansglysnproleuaspalaal 181
|||||
543 ACGGGGGAAGAAAGCGCGGTCACAAACGCAACCGCTGAGCGCGATT 592
|||||
182 glnlin 183
|||||
593 CAGCAG 598

seq_name: qb_pat1:A39567

seq_documentation_block:
LOCUS A39567 1296 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 6 from Patent WO9415644.
ACCESSION A39567
VERSION A39567.1 GI:2295849
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Epenetos,A.A., Spooner,R.A. and Deonaraia,M.
TITLE COMPOUNDS FOR TARGETING
JOURNAL Patent: WO 9415644-A 6 21-JUL-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication GB 2289679 951129.
FEATURES
source
1..1296
1..1296=unidentified"

```

```

/db_xref="taxon:3204"
40..1287
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02450.1"
/db_xref="GI:2295850"
/translation="MKYLLPTAAAGLLILPAQAQYVLOQPGAEILVKGASVKLSC
/translation="MKYLLPTAAAGLLILPAQAQYVLOQPGAEILVKGASVKLSC
KASVLTSTTMMHWVQKQKRGLEWIGRIIDPMSGGTCKNKEFKSKATLTVDKPSSTAV
MOLSLTSEDSAVYTCARDYDSSSEYEDWGGGTGYVSSGGGSGGSGGSGQAV
VTQSSALTSPEGETVLVTCRSTSGAVTTSNANWQEPDHLFTLIGTNAPAGVPA
ARFGSLGIDKKAALTITGAQTEDEALYFCALWISNNHWYGGCTKLTGLGAPAAPAA
DPSKSRKQVSAAPAGITGTWYMQLSSTLVPAQDAGALGTGYBEAIVONASRYLVIG
RYSDSPATDGGGTALGWTVAMKNNYNNRNSHSAITWSSQYVGGAEARINTQMLTSGTIE
ANAKRSTLYGHDPETTKVPSAAS"
BASE COUNT      312 a      342 c      347 g      295 t
ORIGIN
alignment_scores:
    Quality: 140.00      Length: 140
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-589-870-2 x A39567      ..

```

```

Align seg 1/1 to: A39567 from: 1 to: 1296
24 AlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGluValAlaG 40
   |||||
865 GCGACCCCGCCCAAGAACCTCCAAAGCTACAGGTTTCTGCACCGCGAAGCGNG 91
40 yllethmglythmtrptfyrAsnGlnleuclyserthrpheIleValTrpA 57

```

915 TATCACTGGCACCTGGTATACCACTGGGGTTCGACTTTCATTGTGACCG 964
57 IAGIAlAAspGIAlAlALeuThrGIYThrTyGluSerAlAlValGIAsn 73
965 CTGGGCGCGGACGAGGCTCTGACCTGGCGACCTACGAATCTCGGTTGGTAAc 1014
74 AlAgluserArgTyRValLeuThrGIYArgTyRAspSerAlAProAlAth 90
1015 GCAGAAATCCCGCTAGCTAGCTACAGGCGCGGTTAAAGACTCTGCACCTGCCAC 1064
90 IAspGIYSerGIYThrAlAlALeuGIYTPThrValAlAlATrplysAsnAsnT 107
1065 CGATGGCTCTGGTACCGCTCTGGGCTGGACTGTGGCTGGAAAAACAACt 1114
107 YTrArgSnaAlAhISerAlAThrThrTrpSerGIYnTYrValGIYGIY 123
1115 ATGTGATAGCGCACAGCGCACTAGCTGCTGGCAATACGTGGCGGGT 1164
124 AlAglAlAlAArgILeAsnThrGIInTrpLeuLeuThrSerGIYThrThrGI 140
1165 GCGAGAGCTCGTATCAACACTAGGCGGTGTAAATCAATCGGCACTACCGA 1214
140 uAlAsnaAlATrplySerThrLeuValGIYHISAspThrPheThrIYsY 157
1215 AGCGAAATGCATGGAAATCGACACTAGTAGTCATGACACCTTACCAAG 1264
157 AlIYsProSerAlAlAlAser 163
1265 TTAAAGCTTCTGCTGCTAGC 1284

FEATURES	source	base count	origin
LOCUS	A93154	1296 bp	DNA
DEFINITION	Sequence 6 from Patent EP0815872.		
ACCESSION	A93154		
VERSION	A93154.1	GI:6741542	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 1296)		
AUTHORS	Epenetos,A.A. and Deonarain,M.		
TITLE	Compounds for targeting		
JOURNAL	Patent: EP 0815872-A 07-JUN-1998;		
	IMP CANCER RES TECH (GB)		
	Location/Qualifiers		
	1..1296		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
	40..1287		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAB69541.1"		
	/db_xref="gi:6741543"		
	/translation="MKYLLPTAAAGLLLLAOPRANMNOLOOGPAELVYKPGAYKLSGKSGFTFSYMMHWKOPRGRGLEWIRIDPNKGGTKYENKRSKATLLTVKPSSTVAYMOLSLITSDSAVYVCARYDYGSSPYDWGGTIVTVSSGGGSGGGGGGGGSOAVMDSLITSDSPEVTLTTCRSTSGAVTTSYAMVVEKPDHLFTGLIGETNNRAPGVYVDSLESLITTCGATITGAOTDEDAIVFCALWYSNHWVFGGGTKRLTGLIGETNNRAPGVYARRSGSLIDKRAALITGAOTDEDAIVFCALWYSNHWVFGGGTKRLTGLIGETNNRAPGVYDPSGSKAQVYSAAEAGITGTVYNOLGSTFIVTGGADAGALTGYESAAGVNAESRYVLTGRYDASATATGCGTALGWTGYAANKNNYRNHASTTWSGGVYOGAAGARINTOMLTLSGTTTANAWKSTLIGHDTFFPKYVPSAS"		

```
alignment-scores:      length: 140
                    quality: 140.00
                    ratio: 1.000
                    gaps: 0
Percent Similarity: 100.000
Percent Identity: 100.000
```

alignment_block:
US-09-589-870-2 x A93154 ..

Align seg 1/1 to: A93154 from: 1 to: 1296

```
24 ALaasProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCCGCTCCAAGAGACTCCAAGCTCAGGTTTCTGCAGCGGAAGCTGG 914
40 YIleThrGlyThrTrpTyraSngInLeuGlySerThrPheIleValThrA 57
|||||
915 TATCACTGGCACCTGGTATTAACAACACTGGGGTGCACCTTTCATTGTGACCG 964
57 lAGlYAlaSPGlyAlaLeuThrGlyThrTyrgLysSerAlaValGlyAsn 73
|||||
965 CTGGTGGCGAGCGAGCTGTGACTGGCACCTACGAATCTGCGGTGGTAAAC 1014
74 AlAGlUsErArGTyRValLeuThrGlyArGTyRAspSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACCTGGCCGCTTATGACTCTGCACCTGCCAC 1064
90 rAsPGlYsErGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
|||||
1065 CGATGGCTCTGTGTAACCGCTGTGGCTGGACTGTGGCTTGGAACAACT 1114
107 YrArGAsnAlaHisSerAlaThrTrpTrpSerGlyGlnTyRValGlyGly 123
|||||
1115 ATCGTAATGCCACACGCCCACTACGTGTGCTGTGGCAATACGTTGGCGGT 1164
124 AlAGlUAlaArG1eAsnThrGlnTrpLeuLeuThrSerGlyThrG1 140
|||||
1165 GCTGAGCTCGTATCAACACACAGTGGCTGTTAACATCCGGCACACACGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGAATAATGCACACTAGTAGTCATGACACCTTACCAAG 1264
157 aLlYsPrOsErAlaAlaSer 163
|||||
1265 TTAAACCTTGTGCTGTAGC 1284
```

seq_name: gb_pat1:AR082491

seq_documentation_block:

LOCUS AR082491 1296 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 6 from patent US 5973116.

ACCESSION AR082491

VERSION AR082491.1 GI:10009217

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1296)

AUTHORS Epenetos,A,Antoniou, Spooner,R,Anthony and DeonaraIn,M.

TITLE Epenetos,A,Antoniou, Spooner,R,Anthony and DeonaraIn,M.

JOURNAL Patent: US 5973116-A 6 26-OCT-1999;

FEATURES Location/Qualifiers

1..1296

source /organism="Unknown"

BASE COUNT 312 a 342 c 347 g 295 t

ORIGIN

alignment_scores:

Quality: 140.00 Length: 140
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x AR082491 ..

Align seg 1/1 to: AR082491 from: 1 to: 1296

```
24 ALaasProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCCGCTCCAAGAGACTCCAAGCTCAGGTTTCTGCAGCGGAAGCTGG 914
40 YIleThrGlyThrTrpTyraSngInLeuGlySerThrPheIleValThrA 57
|||||
915 TATCACTGGCACCTGGTATTAACAACACTGGGGTGCACCTTTCATTGTGACCG 964
57 lAGlYAlaSPGlyAlaLeuThrGlyThrTyrgLysSerAlaValGlyAsn 73
|||||
965 CTGGTGGCGAGCGAGCTGTGACTGGCACCTACGAATCTGCGGTGGTAAAC 1014
74 AlAGlUsErArGTyRValLeuThrGlyArGTyRAspSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACCTGGCCGCTTATGACTCTGCACCTGCCAC 1064
90 rAsPGlYsErGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
|||||
1065 CGATGGCTCTGTGTAACCGCTGTGGCTGGACTGTGGCTTGGAACAACT 1114
107 YrArGAsnAlaHisSerAlaThrTrpTrpSerGlyGlnTyRValGlyGly 123
|||||
1115 ATCGTAATGCCACACGCCCACTACGTGTGCTGTGGCAATACGTTGGCGGT 1164
124 AlAGlUAlaArG1eAsnThrGlnTrpLeuLeuThrSerGlyThrG1 140
|||||
1165 GCTGAGCTCGTATCAACACACAGTGGCTGTTAACATCCGGCACACACGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGAATAATGCACACTAGTAGTCATGACACCTTACCAAG 1264
157 aLlYsPrOsErAlaAlaSer 163
|||||
1265 TTAAACCTTGTGCTGTAGC 1284
```



```

|||||
1 ATGCCAGATGTCGTGCACACCATCGCCGTTTCCCTGACACGCTGTC 50
17 rtleThraSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
51 GATTCAGCGCAGCGCTCGGCGAGACCCCTCCAAAGACTCGAAGGCCAGG 100
34 a1SerAlaAlaGlnAlaGlyIleThrGlyThrTPTyrAsnGlnLeuGly 50
|||||
101 TCTCGGCGCGCGAGCGGCATCCAGCGCATCGCTGTACAAACGCTCGGC 150
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
151 TCGACCTTCATGTCATCGCGCGCGCGCCAGCGCGCTGACCGGAACCTA 200
67 rgluSerAlaValGlyAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
201 CGAGTCGGCGCGCGCAACGCGCGAGCGCGTACGTCCTGACCGGTCGTT 250
84 yraspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
251 ACGACAGCGCGCGCGCGCACCGGCGAGCGGCGACCGCCCTCGGTTGGAG 300
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
301 GTGGCGCTGGAAGAATACCTACCGCAACGCCCATCCGCGCACACGTCGAG 350
117 rglGlnTyrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
351 CGGCCAGTACGTCGGCGCGCGCGAGCGCGAGTCAACACCCAGTGGCTGC 400
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
401 TGACCTCGGCGCACCGACCGAGCGCAACGCTGGAAGTCCACGCTGTCGCG 450
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
451 CACGACACCTTCACCAAGGTGAAGCGCTCGCGGCTCCATGAGAGCGCGC 500
167 aLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
501 GAAGAAGCGCGCGCTCAACAACGCGCAACCGCGCTCGACCGCTTCAGCAG 549

seq_name: /cgm2_2/gcgdata/geneseq/geneseqn/NA2000.DAT.291073
seq_documentation_block:
ID 291073 standard; DNA; 563 BP.
XX
AC 291073;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomycetes avidinl1 sps gene.
XX
KW plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; sps gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomycetes avidinl1.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PE 30-JUL-1999; 99MO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
DR WPI; 2000-195402/17.

```

```

DR P-PSDB: Y80512.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 84; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomycetes
CC avidinl1 sps gene as an example of the heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 563 BP; 107 A; 213 C; 169 G; 74 T; 0 other;

alignment_scores:
Quality: 183.00 Length: 183
Ratio: 1.000 Gaps: 0
Percent similarity: 100.000 Percent identity: 100.000

alignment_block:
US-09-589-870-2 x 291073 ..

Align seg 1/1 to: 291073 from: 1 to: 563

1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSe 17
|||||
10 ATGCCAGATGTCGTTCGACACCATCGCCGTTTCCCTGACACGCTGTC 59
17 rtleThraSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
|||||
60 GATTCAGCGCAGCGCTCGGCGAGACCCCTCCAAAGACTCGAAGGCCAGG 109
34 a1SerAlaAlaGlnAlaGlyIleThrGlyThrTPTyrAsnGlnLeuGly 50
|||||
110 TCTCGGCGCGCGAGCGGCATCCAGCGCATCGCTGTACAAACGCTCGCG 159
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
160 TCGACCTTCATGTCATCGCGCGCGCGCCAGCGCGCTGACCGGAACCTA 209
67 rgluSerAlaValGlyAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
210 CGAGTCGGCGCGCGCAACGCGCGAGCGCGTACGTCCTGACCGGTCGTT 259
84 yraspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
260 ACGACAGCGCGCGCGCGCACCGGCGAGCGGCGACCGCCCTCGGTTGGAG 309
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
310 GTGGCGCTGGAAGAATACCTACCGCAACGCCCATCCGCGCACACGTCGAG 359
117 rglGlnTyrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
360 CGGCCAGTACGTCGGCGCGCGAGCGGAGTCAACACCCAGTGGCTGC 409
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
410 TGACCTCGGCGCACCGACCGAGCGCATCGGAAGTCCACGCTGTCGCGC 459
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
|||||

```

460 CACGACACCTTCACCAAGGTGAAGCCGTCCGCCCTCCATCGACGCGGC 509
167 alyslsylaaglyValaAsnAsnGlyAsnProleuAspAlaValGlnGln 183
|||||
510 GAAGAAGCGCGCGCTCACACACGCAACCCGCTCGACGCGCTTCACAGCAG 558

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT.T73193

seq_documentation_block:

ID T73193 standard; DNA; 638 BP.

XX AC T73193;

XX DT 27-APR-1998 (first entry)

XX DE Wild-type streptavidin DNA.

XX KM Streptavidin; biotin; anti-interference reagent; detection; mutein;

XX KW avidin; non-specific binding; ss.

XX OS Unidentified.

FH Key Location/Qualifiers
FT CDS 50..601
FT /*tag= a

FT /*product= streptavidin

FT sig_peptide 50..121

FT mat_peptide 122..598

FT /*tag= b
FT /*tag= c
FT /*product= streptavidin

XX DE19637718-A1.

XX PD 02-OCT-1997.

XX PF 16-SEP-1996; 96DE-1037718.

XX PR 01-APR-1996; 96DE-1013053.

XX PA (BOEF) BOEHRINGER MANNHEIM GMBH.

XX PI Brandtetter H, Deger A, Engh R, Kopetzki E, Mueller R;
XX Schmitt U;

XX WP1: 1997-482043/45.

XX DR P-PSDB; W29306.

XX PT Streptavidin and avidin muteins with reduced binding affinity for
XX biotin - useful for reducing interference from nonspecific binding
XX in assays

XX PS Disclosure; Page 16-17; 26pp; German.

XX This DNA sequence encodes a streptavidin which is used in a novel method
XX of reducing interference from non-specific binding in assays. Muteins
XX constructed from a core streptavidin or avidin sequence are selected that
XX differ from the native polypeptide by at least one amino acid and have a
XX binding affinity for biotin of less than 1010 l/mole. The biotin-bindable
XX polypeptide may be present as a polymeric conjugate, e.g. with another
XX polypeptide or protein, especially bovine serum albumin. These muteins
XX are used as anti-interference reagents for reducing and/or avoiding
XX nonspecific interactions in a process for detecting an analyte. In
XX particular, they are used in assays where the streptavidin/avidin-biotin
XX specific binding pair is involved for qualitative and/or quantitative
XX determination of an analyte in a test sample, e.g. a heterogeneous
XX immunassay or a hybridisation assay. Despite having a lower binding
XX affinity for biotin, the muteins have high immunological cross-reactivity
XX with native streptavidin and avidin.

SQ Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

alignment_scores:
Quality: 183.00 Length: 183
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x T73193 ..

Align seg 1/1 to: T73193 from: 1 to: 638

1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrValSe 17
|||||
50 ATGCCCAAGATCGTCTGCACCAATCGCGCTTCCCTGACCAAGCTTC 99
17 rLleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
|||||
100 GATTACGGCCAGCGCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCACAG 149
34 aLSeAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
|||||
150 TCTCGCGCGCGCGCGCATCACGCGCACCTGGTGTACACACAGCTCGCG 199
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
200 TCGACCTTCATCGTGAACCGCGCGCGCGCGCGCGCTGACCGGAACCTA 249
67 rGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||
250 CGAGTCGGCGCGTGGCGACGCGGAGAGCGCGTACGTCGACCGGCTGCTT 299
84 yRAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
|||||
300 ACGACAGCG 349
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
|||||
350 GTGGCCTGGAGAAATTAATACGCAACCGCCACTCCGCGACACAGTGGAG 399
117 rGlyGlnTyrValGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
|||||
400 CGGCGCATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 449
134 euThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCCGCGCACACGACGCGCAACGCTCGAAGTCCACGCTGTCGCG 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
|||||
500 CACGACACCTTCACCAAGGTGAAGCGTCCGCGCGCTCCATCGACGCGCG 549
167 alyslsylaaglyValaAsnAsnGlyAsnProleuAspAlaValGlnGln 183
|||||
550 GAAGAAGCGCGCGCTCACACACGCAACCCGCTCGACGCGCTTCACAGCAG 598

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT.V34714

seq_documentation_block:

ID V34714 standard; DNA; 638 BP.

XX AC V34714;

XX DT 27-AUG-1998 (first entry)

XX DE S. avidinii streptavidin cDNA.

XX KM Streptavidin; ligand; binding affinity; mutant; isolation;

XX KW purification; recover; immobilise; ss.

XX OS Streptomyces avidinii.

FH Key Location/Qualifiers
FT CDS 50..601
FT /*tag= a

```

FT      sig_peptide      50..121
FT      /*tag= b
FT      mat_peptide      122..598
FT      /*tag= c
FT      /product= Streptavidin
XX
XX      EP835934-A2.
XX      15-APR-1998.
XX
XX      09-OCT-1997; 97EP-0117504.
XX
XX      10-OCT-1996; 96DE-1041876.
XX
XX      (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX      Skeria A, Voss S;
XX      PI
XX      WPI: 1998-218868/20.
XX      P-PSDB; W59216.
XX
XX      Streptavidin mutants with higher binding affinity for peptide
XX      ligands - have mutation in amino acid region 44-53, used to isolate,
XX      purify or determine fusion proteins including these ligands
XX
XX      Disclosure; Page -: 21pp; German.

```

```

XX      This sequence encodes a wild-type streptavidin protein isolated from
XX      Streptomyces avidinii. This sequence is used to produce mutants which
XX      are used in a method to assay the binding affinity of streptavidin
XX      mutants. These mutants have a mutation within the amino acid (aa) region
XX      44-53 of the wild-type protein show a higher binding affinity than the
XX      wild-type for peptide ligands that include the sequence of formula
XX      TRP-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly,
XX      or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can
XX      be used to isolate, purify and determine proteins or to determine/recover
XX      substances that contain streptavidin-binding groups. Such compounds may
XX      also be used to immobilise fusions on microtitre plates, microbeads or
XX      sensor chips.
XX      NOTE: This sequence does not appear in the specification but is used to
XX      make the mutant streptavidin sequence represented in V34715 and V34716.
XX
XX      Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

```

```

XX      Alignment_scores:
XX      Quality: 183.00      Length: 183
XX      Ratio: 1.000      Gaps: 0
XX      Percent Similarity: 100.000      Percent Identity: 100.000

```

alignment_block:

US-09-589-870-2 x V34714 ..

Align seg 1/1 to: V34714 from: 1 to: 638

```

1 MetArgLysIleValAlaIleAlaValSerLeuThrThrValSe 17
|||||
50 ATGCCAAGATGCTGTGCAGCCATGCCGCTTCTCCAGCACGGCTCC 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
|||||
100 GATTACGGCCAGCGCTTGGCGCAGACCTCCAGAGACTCGAAGGCCACG 149
|||||
34 aLserAlaAlaGluAlaGlyIleThrGlyThrTriPtyrAsnGlnLeuGly 50
|||||
150 TCTCGCGCGCGAGCGCGCATCACCGCAGACTGCTACACACAGCTCGGC 199
|||||
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThy 67
|||||
200 TCGACCTTCATCGTGCACCGCGCGCCGACGCGCTGACCGGAACCTA 249
|||||
67 rGluSerAlaValAlaGluAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||

```

```

250 CGAGTCGCGCTGCGCAACGCCGAGACCGCTACGTCGTGACCGGTCGT 299
84 yRAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuLysTriPthr 100
|||||
300 ACGACAGCGCCCGCCGACCGAGCGAGCGGCGCCCTCGGTTGGAGC 349
|||||
101 ValAlaThrLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
|||||
350 GTGGCTCGAAGATACACTACCGCAACGCCCACTCCGCGACACAGTGGAG 399
|||||
117 rGlyGlnTyrValGlyGlyAlaGluAlaArgTlleAsnThrGlnTrpLeuL 134
|||||
400 CGGCAGTACGTGCGCGCGCGAGCGAGATCAACACCGAGTGGCTGC 449
|||||
134 eUrThrSerGlyThrThrGluAlaAsnAlaTriPtyrSerThrLeuValGly 150
|||||
450 TGACCTCCGCGCACACGAGCGCAACGCTCGAAGTCCACGCTGGTGGCG 499
|||||
151 HisAspThrPheThrLysValLysProSerAlaIleAlaSerLleAspAla 167
|||||
500 CAGGACACCTTCACCAAGGTGAAGCGGTCCGCGCTCATCGACGCGGC 549
|||||
167 aLysLysAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 GAAGAAGCGCGCGCTCAACACGCAACGCCGCTCGACGCGCTTCAGCAG 598

```

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:Z49867

seq_documentation_block:

ID Z49867 standard; cDNA; 638 BP.

XX Z49867;

DT 25-APR-2000 (first entry)

DE Streptavidin gene for recombinant pART27 vector.

KW potato proteinase inhibitor-II; PPI-II; streptavidin; worm;

KW insect; plant-noxious protein; pest resistance; moth; insect; weevil;

KW grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;

KW insecticidal; ss.

XX OS Synthetic.

FH Key Location/Qualifiers

FT CDS 50..601

FT /*tag= a

FT /product= "Streptavidin protein"

FT sig_peptide 50..121

FT /*tag= b

FT mat_peptide 122..598

FT /*tag= c

XX WO200004049-A1.

XX 27-JAN-2000.

XX 15-JUL-1999; 99WO-N200110.

XX 15-JUL-1998; 98NZ-0331002.

XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.

XX Christeller JT, Sutherland PW, Murray C, Markwick NP, Phillip BA;

XX Malone IA, Burgess EPJ;

XX WPI: 2000-171244/15.

XX P-PSDB; Y44701.

XX New chimeric polypeptide and composition comprising the polypeptide

XX useful for conferring pest resistance on plants

XX Disclosure; Fig 12; 11pp; English.


```

|||||
530 GATTAGGCGCAGGCTTGGCAGACCCCTCCAAAGCACTCGAAGCCGACG 579
34 aISerAlaIaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
|||||
580 TCTCCGCCCCCGAGGCGGCATCACCGGCACCTGGTACCAACGAGCTCGGC 629
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
630 TCGACCTTCATCTGACCGCGGGCGCGAGCGGGCCCTGACCGGAACCTGA 679
67 rGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||
680 CGAGTGGCGCGTCGGCAACGCGAGAGCGCGTACGCTCCGACGGGTCGT 729
84 yraspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
|||||
730 ACGACAGCGCCCGCGCACCGACGCGAGCGGCGCACCCCTCGGTTGAGCG 779
101 ValAlaIleThrPlysAsnAspTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
|||||
780 GTGGCCTGGAGAGATTAATCAACGCAAGCCCACTCCGCAACCACTGGAG 829
117 rGlyGlnTyrValAlaGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
|||||
830 CGGCGCATGACGTGCGGCGCGCGAGCGAGATCAACACCCAGTGGCTCG 879
134 eutHserGlyThrThrGluAlaAsnAlaIleThrPlysSerThrLeuValGly 150
|||||
880 TGACCTCCGCGCACACCGAGCCCAACGCTGGAAGTCCACGCGTGTGCGC 929
151 HisAspThrPheThrIleValAlaLysProSerAlaAlaSerIleAspAlaI 167
|||||
930 CACGACACCTTCAACCAAGGTGAAGCCGTCCGCGCCTCCATCGACCGCGC 979
167 aLysIlyAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValAlaGlnGln 183
|||||
980 GAAGAAGGCGCGGCGTCAACACGCGCAACCCGCGTCAACGCGCTTCACAG 1028
seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:291074
seq_documentation_block:
ID 291074 standard; DNA: 564 BP.
AC 291074;
AT 06-JUN-2000 (first entry)
DT 06-JUN-2000 (first entry)
DE Streptomyces avidinii mst gene.
KW Plant somatic tissue degeneration; plant essential factor; depletion;
viability; mst gene; plant development; plant morphology; flower;
fruit plant; ds.
XX Streptomyces avidinii.
OS
PN WO200007427-A2.
PD 17-FEB-2000.
PF 30-JUL-1999; 99WO-IL00420.
PR 03-AUG-1998; 98IL-0125632.
PA (AGRI-) AGRIC RES ORG.
PI Kapulnik Y, Ginzberg I;
DR MPI: 2000-195402/17.
P-PSDB: Y80513.
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such

```

```

PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX Examples; Page 85; 91pp; English.
PS
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii mst gene as an example of a heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 564 BP; 109 A; 213 C; 169 G; 73 T; 0 other;

alignment_scores:
      quality: 182.00      length: 182
      ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 291074 ..
Align seg 1/1 to: 291074 from: 1 to: 564

2 ArgYIlaValValAlaIleAlaIleAlaValSerLeuThrThrValSerIle 18
|||||
14 CGAAGATGTCGTCCTGACGCCATCGCCGTTCCCTGACCAACGCTCGAT 63
18 eThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValS 35
|||||
64 TACGGCCAGCGCTTGGCGACAGCCCTCCAAAGACTCGAAGGCCACAGCT 113
35 eRAlaIaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGlySer 51
|||||
114 CGGCCCGCGAGGCGGCATCACCGGCACCTGTGTACCAACGCTCGGCTCG 163
52 ThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTrpG 68
|||||
164 ACCCTTCATCTGACCGCGGGCGCGAGCGGCCCTGTACCGGAACCTACGA 213
68 uSerAlaValAlaGlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyr 85
|||||
214 GTGCGCGCTCGGCAACGCGAGAGCGCGTACGTCCTGACCGGTCGTAG 263
85 spSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThrVal 101
|||||
264 ACAGCGCCCGCGCACCGAGCGGACCGCCCTCGGTTGAGCGGTG 313
102 AlaTrpPlysAsnAspTyrArgAsnAlaHisSerAlaThrThrTrpSerG 118
|||||
314 GCCTGGAAGATTAATCAACGCAAGCCCACTCCGCAACCACTGGAGCGG 363
118 yGlnTyrValAlaGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuT 135
|||||
364 CCAGTACGTGCGGCGCGCGAGCGAGATCAACACCAACCGTGGCTCGTA 413
135 hrSerGlyThrThrGluAlaAsnAlaIleThrPlysSerThrLeuValGlyHis 151
|||||
414 CCTCGGCACACCAACGAGCGCCCTGGAAGTCCACGCTGTGCGGCCAC 463
152 AspThrPheThrIleValLysProSerAlaAlaSerIleAspAlaIleAla 168
|||||
464 GACACCTTCAACCAAGGTGAAGCGGTCCGCGCCTCCATCGACGCGGCA 513
168 sIlyAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValAlaGlnGln
|||||

```

514 GAAGCCGCGCTCAACACGACCCGCTCGACCGCTTCAGACG 559
seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:291076
seq_documentation_block:
ID 291076 standard; DNA: 495 BP.
XX 291076;
AC
XX 06-JUN-2000 (first entry)
DT
XX Streptomyces avidinii mprot gene.
DE
XX Plant somatic tissue degeneration: plant essential factor: depletion;
KW viability; mprot gene; plant development; plant morphology; flower;
KM fruit plant; ds.
XX
OS Streptomyces avidinii.
XX W0200007427-A2.
PN
XX 17-FEB-2000.
PD
XX 30-JUL-1999; 99MO-I100420.
PE
XX 03-AUG-1998; 98IL-0125632.
PR
XX (AGRI-) AGRIC RES ORG.
PA
XX Kapulnik Y, Glinzberg I;
PI
XX WPI: 2000-195402/17.
DR P-PSDB; Y80515.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
as decreasing the number of flowers present to increase the number of
PT fruit -
XX
XX Examples: Page 86-87; 91pp; English.
PS
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii mprot gene as an example of a heterologous gene introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 495 BP; 98 A; 188 C; 151 G; 58 T; 0 other:

Alignment_scores:
Quality: 160.00 Length: 160
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-589-870-2 x 291076 ..

Align seg 1/1 to: 291076 from: 1 to: 495

24 AIAAAPPROSerLyAspSerLyAGlnValSerAlaIaGluAaG1 40
|||||
11 GGTGACCCCTCCAGAGACTCGAAGGCCAGGTCTCGCGCCGAGGCCG 60
40 ylleThrGlyThrTPTyFAnglInLeuGlySerThrPheIleValThra 57
|||||

61 CATCACCGGACACCTGGTATCACACAGCTCGCTGACCTTCATCTGACCG 110
57 IagLYAlaSpCIyAlaLeuThrGlyThrTyrGluSerAlaValGlyAsn 73
|||||
111 CGGGGCGCGGCGCGCCCTGACCGGAAGCTACGAGTCGGCGCTGGCAAC 160
74 AlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaTh 90
|||||
161 GCCGAGAGCGGCTACGCTCTGACCGGTCTGTCAGACAGGCCCGGCCGAC 210
90 rAspGlySerGlyThrAlaLeuGlyTyrThrValAlaTyrPlyAsnAsnT 107
|||||
211 CGACGGCAGCGGACCGCCCTCGTTGAGCGGTGGCTGGAAGAAATAACT 260
107 yTArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGly 123
|||||
261 ACCGCAACGCCACCTCCGCGGACCGTGGAGCGGCAGATCACTCGCGCGC 310
124 AlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
311 GCCGAGCGGAGATCAACACCCAGTGGCTGCTGACCTCCGCGACACCGCA 360
140 uAlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
361 GCCCAACGCGCTGGAAGTCCACGCTGGTGGCCACGACACCTTCACCAAG 410
157 ALLySPROSerAlaAlaSerIleAspAlaAlaLysAlaGlyValaAsn 173
|||||
411 TGAAGCCGTCCGCGCTTCATCGACGCGGCGAAGACGCGCGCTCAAC 460
174 AsnGlyAsnProLeuAspAlaValGlnGln 183
461 AACGGCAACCCGCTCGACGCCGTTCAACGAG 490

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT:070659
seq_documentation_block:
ID 070659 standard; DNA: 1356 BP.
XX 070659;
XX
XX 26-MAR-1995 (first entry)
DT
XX SCFV PRAS108 and PRAS112 DNA.
DE
XX Amplification: single chain variable region fusion protein; PCR; ss.
KW
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT CDS 40..1348
FT /*tag= a
XX
XX W09415644-A.
FN
XX
XX 21-JUL-1994.
PD
XX
XX 17-JAN-1994; 94MO-GB00087.
PE
XX
XX 15-JAN-1993; 93GB-0000686.
PR
XX
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA
XX
XX Deonarain M, Epenetos AA, Spooner RA;
PI
XX
XX WPI: 1994-248907/30.
DR
XX
XX P-PSDB; R56483.
DR
XX
XX New cpds. comprising a targeting portion and a cytotoxic portion
PT - used esp. for treating mammals for destroying target cells,
PT partic. tumour cells
XX
XX Claim 36; Fig 9; 114pp; English.
PS

XX The sequence is that of the ScFv pRAS108 and pRAS112 between HindIII and
 CC EcoRI site obtd. by PCR.
 CC See also Q70652-69.
 CC

SQ Sequence 1356 BP; 332 A; 359 C; 359 G; 306 T; 0 other;

alignment_scores:
 Quality: 160.00 Length: 160
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x Q70659 ..

Align seg 1/1 to: Q70659 from: 1 to: 1356

24 Alaasproserlyaspserysalaaglnvalseralaalaglualag1 40
 ||||||||||||||||||||||||||||||||||||||||||||||||
 865 GCAGACCGCTCCAAAGACTCCAAAGCTCAGGTTTCTGCAGCGGAAGCTGG 914
 40 yllethrglythrtrpyrasnglnleuglyserthrpheillevalthra 57
 ||||||||||||||||||||||||||||||||||||||||||||||||
 915 TATCACTGGACCTGCTATTAACCAACTGGGGCTGACTTTCATTCTGACCG 964
 57 laglyalaspolyalaleuthrglythrpyrgluseralavalglyasn 73
 ||||||||||||||||||||||||||||||||||||||||||||||||
 965 CTGGGCGGAGGAGCTCTGACTGGACCTGACGAATCTGGGGTTGGTAAAC 1014
 74 Alaiguseryarqyvalleuthrglythrpyraspseralaprolath 90
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1015 GCAGAAATCCCGGTACTGACTGCGCGGTATGACTCTGCGACCGCCAC 1064
 90 rAspglyserglythralaleuglytrpThrValAlatrplysasnasnt 107
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1065 CGATGGCTCTGTGTACCGCTGTGGCTGACTGTGTGGAATAAACAACT 1114
 107 yrrarAsnalhisseralathrtrpserglylnltyrvalglyly 123
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1115 ATCGTAATGCGACAGCGCCACTGAGTGTGCGCAATACCTTGGCGGT 1164
 124 AlaAlaualarqileasnThrglntrpleuLeuthrSerglyThrThrG1 140
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1165 GGTGAGGCTCGTATCACACACTGAGTGGCTTAACATCGGACATACGGA 1214
 140 uAlaAsnalatrplyserThrleuValGlyHisasphrPheThrLysV 157
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1215 AGCGAATGCATGGAATCGACACTAGTAGTGTATGACACCTTACCAAG 1264
 157 allyproserAlaAlaserlleasppAlaAlaValysAlaGlyValasn 173
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1265 TTAAACCTTCTGCTGCTAGCATGATGCTGCCAAGAAAGACGAGGCTTAAC 1314
 174 AsnglyAsnProLeuAspAlaValGlnGln 183
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1315 AACGGTAACCTCTAGACGCTTTTACGACA 1344

seq_name: /cgn2_2/gcndata/geneseq/geneseqn/NA2000.DAT:Z91075

seq_documentation_block:

ID Z91075 standard; DNA; 492 BP.

XX Z91075;

DT 06-JUN-2000 (first entry)

DE Streptomyces avidinii prot gene.

XX Plant somatic tissue degeneration; plant essential factor; depletion;
 KW viability; prot gene; plant development; plant morphology; flower;
 KM fruit plant; ds.

XX

OS Streptomyces avidinii.

XX PN W0200007427-A2.

XX PD 17-FEB-2000.

XX PF 30-JUL-1999; 99W0-IL00420.

XX PR 03-AUG-1998; 98IL-0125632.

XX PA (AGRI-) AGRIC RES ORG.

XX PI Kapulnik Y, Ginzberg I;

XX DR WPI: 2000-195402/17.

XX DR P-PSDB; Y80514.

XX PT Degeneration of somatic plant tissue by expression of a heterologous
 PT protein, useful for controlling plant development and morphology, such
 PT as decreasing the number of flowers present to increase the number of
 PT fruit -

XX Examples; Page 86; 91pp; English.

XX The invention relates to a method of effecting degeneration of a somatic
 CC plant tissue by expressing a heterologous protein capable of binding a
 CC plant essential factor (PEF), in somatic plant tissue cells, where
 CC heterologous protein expression causes depletion of the PEF so the plant
 CC viability is maintained, while simultaneous degeneration of the somatic
 CC plant tissue is effected. This sequence represents the Streptomyces
 CC avidinii prot gene as an example of a heterologous gene introduced into
 CC the plants. The methods can provide for the selective and optionally
 CC reversible cell degeneration in somatic plant tissue. They can be used
 CC for artificially controlling plant development and morphology. They can
 CC be used e.g. to decrease the number of flowers in fruit producing plants
 CC so as to increase the number of fruits which reach maturity.

SQ Sequence 492 BP; 95 A; 188 C; 151 G; 58 T; 0 other;

alignment_scores:
 Quality: 159.00 Length: 159
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x Z91075 ..

Align seg 1/1 to: Z91075 from: 1 to: 492

25 Aspproserlyaspserysalaaglnvalseralaalaglualagly11 41
 ||||||||||||||||||||||||||||||||||||||||||||||||
 11 GACCCCTCCAAAGACTCGAAGGCCAGGCTCTGGCCCGGAGGCGGAT 60
 41 eThrglythrtrpyrasnglnleuglyserthrpheillevalthralag 58
 ||||||||||||||||||||||||||||||||||||||||||||||||
 61 CACGCGACACTGCTAACAACACTGCGCTCGACCTTCATCGAGACGGGG 110
 58 lYAlaAspGlyAlaLeuthrglythrTrpGluSerAlaValGlyAsnAla 74
 ||||||||||||||||||||||||||||||||||||||||||||||||
 111 GCGCCGACGCGCCCTGACCGGAACCTACGAGTCCGCGCGGCAACGCC 160
 75 GluserArqyValleuthrglythrpyraspseralaprolathras 91
 ||||||||||||||||||||||||||||||||||||||||||||||||
 161 GAGAGCCGCTACGCTCGACCGTGTTCAGACGACGCGCCGACCGA 210
 91 pGlySerglyThralaleuglyTrpThrValAlatrplysasnasntYrA 108
 ||||||||||||||||||||||||||||||||||||||||||||||||
 211 CGGACGCGGACCGCCCTCGGTGGACGCTGCGAAGAAATACACTACC 260
 108 rGAsnalAHisserAlaThrTrpSerglyGlnTrpValGlyGlyAla 124
 ||||||||||||||||||||||||||||||||||||||||||||||||
 261 GCAAGCCCACTCCGCGACACAGTGGAGGGCGCACTAGCTGCGGCGGCC 310


```
125 GUAAlaArgIleasnThrGlnTrpLeuThrSerGlyThrThrGlnA 141
|||||
311 GAGCGCAGATCATACACCCAGTGGCTGCTGACCTCGGCAACCCAGGC 360
141 aasnalaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
361 CAACGCGTGAAGTCCAGCTGTGTCGCCACGACCTTCACCAAGTGA 410
158 ysProSerAlaAlaSerIleAspAlaLysLysAlaGlyValAsnAsn 174
411 AGCCGTCGCCCGCTTCATCGACGCGCGAAGAGCGCGCTCAACAAC 460
175 GlyAsnProLeuAspAlaValGlnGln 183
461 GGCACCCGCTGCACGCCGCTTCAGCAG 487

seq_name: /cgn2_2/gcgdata/geneseq/geneseq/NA1988.DAT: N80186

seq_documentation_block:
ID N80186 standard; DNA: 498 BP.
AC N80186;
XX
XX 13-OCT-1990 (first entry)
XX
XX Encodes biosynthetic protein with streptavidin trailer.
DE
XX Biosynthetic Antibody Binding site (BABS); variable domain;
KW anti-CPA monoclonal antibody; streptavidin; ss.
XX
XX Synthetic.
XX
XX W08809344-A.
XX
XX 01-DEC-1988.
XX
XX 19-MAY-1988; 88WO-US01737.
XX
XX 21-MAY-1987; 87US-0052800.
XX
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX
XX Huston JS, Oppermann H;
XX
XX WPI: 1988-353928/49.
XX
XX P-PSDB; P80160.
XX
XX Recombinant multifunctional protein -
XX having an antibody binding site and a sequence for biological activity,
XX ion sequestering or binding to a solid support.
XX
XX Disclosure; ; 115pp; English.
XX
XX Encodes multi-functional biosynthetic protein comprising single
XX chain BABS and streptavidin protein trailer linked via a spacer
XX sequence.
XX See also N80171-N80192.
XX
XX Sequence 498 BP; 117 A; 137 C; 133 G; 111 T; 0 other;

Alignment_scores:
Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-589-870-2 x N80186 ..

Align seg 1/1 to: N80186 from: 1 to: 498

25 AspProSerLysAspSerLysAlaGlnValSerAlaLysAlaGlyVal 41
```

```
|||||
13 GAGCCGTCACAGGACTCCAAAGCTCAGGTTTGTGTCGCCAAGCTGTA 62
41 eThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAla 58
63 CACTGCGACCTGGTATACCACTGGGGTTCGACTTTCATGTTGACCGCT 112
58 lValAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAla 74
113 GTGCGGACGAGCTGTGACTGGCAGCCTACGAATCTGCGGTGGTAACGA 162
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
163 GATCCCGCTAGCTACTGACTGCGCGTTATGACTCGACCTCCACCGCA 212
91 pGlySerGlyThrAlaLeuGlyTyrThrValAlaTrpLysAsnAsnTyr 108
213 TGCGCTGTGACCGCTCTGGGCTGGACTGTGGCTTGAAAAACACTATC 262
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTrpValGlyValAla 124
263 GTATATCGCACAGCGCCACTACGTGGTCCCAATACGTTGGCGGTGCT 312
125 GUAAlaArgIleasnThrGlnTrpLeuThrSerGlyThrThrGlnA 141
313 GAGCGCTGATATCAACACTCAGTGGCTGTATACATCCGGCACTCCAGAC 362
141 aasnalaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
363 GAATGCATGGAATCGACACTAGTGCATGACACCTTACCAAAAGTTA 412
158 ysProSerAlaAlaSerIleAspAlaLysLysAlaGlyValAsnAsn 174
413 AGCCTTCTGCTGCTAGCATTTGATGTCGCAAGAACGACGTTAAACAAC 462
175 GlyAsnProLeuAspAlaValGlnGln 183
463 GGTAAACCTCTAGACGCTGTTCAGCAA 489

seq_name: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT: T75491

seq_documentation_block:
ID T75491 standard; DNA: 498 BP.
XX
XX T75491;
XX
XX 27-FEB-1998 (first entry)
XX
XX DNA for streptavidin.
XX
XX S-layer; sbs-A; vaccine; adjuvant; carrier; hybridisation assay;
XX molecular spinning nozzle; molecular laser; streptavidin; ds.
XX
XX Streptomyces avidinII.
XX
XX DE19603649-A1.
XX
XX 07-AUG-1997.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX (LOBIT/) LOBITZ W.
XX (SLEY/) SLEYTR U.
XX
XX Kuen B, Lubitz W, Sleytr U;
XX
XX WPI: 1997-394558/37.
XX
XX Preparation of S-layer proteins by expressing sbs-A gene in Gram
XX negative bacterium - or new sbs-B gene in any host, also new
XX recombinant proteins containing heterologous inserts, e.g.
PT
```

PT epitope(s), useful as vaccines and adjuvants

XX
PS Example 8; Page 26; 31pp; German.

XX
CC S-layer structures comprising the novel *Bacillus stearothermophilus*
CC PV72 S-layer protein, sbs-A, can be used as vaccines or adjuvants,
CC particularly when they include a bacterial ghost that may contain
CC additional epitopes in its membrane. Other uses of recombinant
CC sbs-A, depending on the nature of the inserted peptide, are as an
CC universal carrier for biotinylated reactants for use in
CC immunological or hybridisation assays (the insert is streptavidin,
CC preferably encoded by the present sequence), to induce immune
CC responses (epitopes), as a reagent for removing cytokine or toxin
CC from serum (antigenic epitopes), as a molecular spinning nozzle
CC (polyhydroxybutyrate synthase) and as a molecular laser
CC (luciferase).

XX
SQ Sequence 498 BP; 120 A; 137 C; 131 G; 110 T; 0 other;

alignment_scores:

Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x T75491 ..

Align seg 1/1 to: T75491 from: 1 to: 498

25 AASPProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGlyI1 41
|||||
7 GAACCGTCCAAAGACTCCAAAGCTCAGGTTTCGACGCCGAAGCTGTAT 56
41 ethrGlyThrTTPyrAsnGlnLeuGlySerThrPheIleValThAlag 58
|||||
57 CACTGGCACCTGGTATMACCACTGGGGTGCACCTTCATGTTGACCGCTG 106
58 lYAlaAspGlyAlaLeuThrGlyThrTyrGlySerAlaValGlyAsnAla 74
|||||
107 TGCGCGACGAGACTCTGACTGGCACCCTACGAATCTGGGTGTAAACGA 156
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
|||||
157 GAATCCCGCTACGTACTGACTGGCGCTGTATGACTCGACCTGCCACGA 206
91 pGlySerGlyThrAlaLeuGlyTTPThrValAlaTrpLysAsnAsnTyrA 108
|||||
207 TGGCTCTGGTACCGCTGTGGCTGGACTGTGGCTTGGAACCAACTATC 256
108 rGAsnAlaHisSerAlaThrTrpTTPSerGlyGlnTyrValGlyValA 124
|||||
257 GTAAATGGCAGACGCGCCTACGTGCTGGCCAAATACCTTGGCGGTGCT 306
125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGlnAl 141
|||||
307 GAGGCGCTGATCAACACGCTGAGTGTACATCCGCGACATACCGAAGC 356
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
|||||
357 GAATGGATGGAATGCACACTAGTAGTCATGACACCTTACCAAGTAA 406
158 ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
407 ACCCTTCTCTCTCTAGCATGTGCTGCGCAAGAAAGCAGCGCTAAACAC 456
175 GlYAsnProLeuAspAlaValGlnGln 183
|||||
457 GGAAACCCCTCTAGACGCTGTCAGCAA 483

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1989.DAT.N90755

seq_documentation_block:

ID N90755 standard; DNA; 507 BP.
XX
AC N90755;
XX
DT 04-JUN-1990 (first entry)
XX
DE Sequence of streptavidin synthetic gene.
XX
KW Streptavidin; synthetic gene; Streptomyces avidinii;
KW biotin; chimeric gene
XX
FH Key
FT CDS
FT
FT Location/Qualifiers
FT
FT 10..492
FT /tag= a
FT 2..46
FT
FT /tag= b
FT /label=BB214
FT complement (6..53)
FT
FT /tag= c
FT /label=BB215
FT 47..86
FT /tag= d
FT /label=BB216
FT complement (54..93)
FT
FT /tag= e
FT /label=BB217
FT 87..127
FT /tag= f
FT /label=BB218
FT complement (94..134)
FT
FT /tag= g
FT /label=BB219
FT 128..171
FT /tag= h
FT /label=BB220
FT complement (135..178)
FT
FT /tag= i
FT /label=BB221
FT 172..213
FT /tag= j
FT /label=BB222
FT complement (179..220)
FT
FT /tag= k
FT /label=BB223
FT 214..256
FT /tag= l
FT /label=BB224
FT complement (221..268)
FT
FT /tag= m
FT /label=BB225
FT 257..295
FT /tag= n
FT /label=BB226
FT complement (264..302)
FT
FT /tag= o
FT /label=BB227
FT 296..338
FT /tag= p
FT /label=BB228
FT complement (303..345)
FT
FT /tag= q
FT /label=BB229
FT 399..380
FT /tag= r
FT /label=BB230
FT complement (346..387)
FT
FT /tag= s
FT /label=BB231
FT 381..423
FT /tag= t
FT /label=BB232
FT complement (390..431)
FT
FT /tag= u
FT misc_feature

```
FT      /label-BB233
FT      misc_feature 424..465
FT      /tag= v
FT      /label-BB234
FT      misc_feature /label-BB234 complement (432..472)
FT      /tag= w
FT      misc_feature /label-BB235
FT      466..502
FT      /tag= x
FT      misc_feature /label-BB236
FT      complement (473..506)
FT      /tag= y
FT      /label-BB237
XX
XX      WO8903422-A.
XX      20-APR-1989.
XX
XX      07-OCT-1988; 88WO-GB00831.
XX
XX      08-OCT-1987; 87GB-0023661.
XX
XX      (BRBI-) BRIT BIO-TECHN LTD.
XX
XX      Edwards RM;
XX
XX      MPI: 1989-130040/17.
XX      P-PSDB; P93531.
XX
XX      DNA sequence encoding streptavidin and vector -
XX      comprising hybrid gene encoding fusion protein with
XX      biotin-binding activity
XX
XX      Fig 2; page 2/5; 22pp; English.
XX
XX      It is based upon the amino acid sequence of natural mature streptavidin
XX      (P93530). It is constructed from 24 oligomers (BB214-BB237) (see
XX      Features Table) which were synthesised by automated phosphoramidite
XX      chemistry. Complementary oligomers were annealed and the pairs ligated
XX      using T4 DNA ligase. The ligated product was sepd. and used to transform
XX      HB87. Plasmid DNA was isolated and positive clones selected. It has
XX      many restriction sites. Fragments of it are also claimed. The genetic
XX      construct is pref. a vector and comprises a chimeric gene composing all
XX      of a fragment of it fused to any other DNA sequence to result in a
XX      sequence encoding a hybrid protein with biotin-binding activity.
XX
XX      Sequence 507 BP; 125 A; 137 C; 130 G; 115 T; 0 other;

alignment_scores:
      Quality: 159.00      Length: 159
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x N90755 ..

Align seg 1/1 to: N90755 from: 1 to: 507

25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGlyI1 41
|||||
13 GACCGCTCAAGAGACTCCAAAGCTCAGTTTCTGCAGCCGAGCTGGTAT 62
|||||
41 ePhrGlyThrTPrTyraNgInLeuGlySerThrPheIleValThrAlaG 58
|||||
63 CACTGGCAGCTGGTATTAACCACTGGGGTGCAGCTTTCATTGTGACCGCTG 112
|||||
58 LysAlaAspGlyAlaLeuThrGlyThrTyrgLysSerAlaValGlyAsnAla 74
|||||
113 GTGGGGAGGAGAGCTGTGACCTGACCTACGAAATCTGGGTTGGTAACGCA 162
|||||
75 GluSerArgTyValLeuThrGlyArgTyraSpSerAlaProAlaThrAs 91
|||||
```

```
163 GAATCCCGCTACGACTGACTGGCCGTTATGATGACTGTGCACCTGCCACCGA 212
91 pGlySerGlyThrAlaLeuGlyTTPThrValAlaTPrLysAsnAsnTyra 108
|||||
213 TGCGTCTGGTACCGCTCGCTGGCTGGAGCTGTGGCTTGAACAAACATATC 262
|||||
108 rGAsnAlaHisSerAlaThrTPrSerGlyGlnTyValGlyValAla 124
|||||
263 GTAAATGCCGACAGCGCCCACTACGTGCTGTGCCAATACGTTGGCGGTGCT 312
|||||
125 GluAlaArgIleAsnThrGlnTPrLeuLeuThrSerGlyThrThrGluAl 141
|||||
313 GAGCGCTGTATCAACACTCAAGTGGCTGTAAACATCCGCACTACCGAAGC 362
|||||
141 aAsnAlaTPrLysSerThrLeuValGlyHisAspThrPheThrIysValL 158
|||||
363 GAATGCATGGAAATCGACACTAGTAGGTCATGACACCTTACCAAGTGA 412
|||||
158 YSPProSerAlaAlaSerIleAspAlaIleLysLysAlaGlyValAsnAsn 174
|||||
413 AGCCTTCTGCTGCTAGCAATGTGATCTGCCAAGAAAGCAGCGGTAAACAC 462
|||||
175 GlyAsnProLeuAspAlaValGlnGln 183
|||||
463 GGTAAACCTCTAGACGCTGTTCACGCA 489
|||||

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA1993.DAT:Q53412

seq_documentation_block:
ID Q53412 standard; DNA; 552 BP.
XX
XX      Q53412;
XX
XX      27-JUN-1994 (first entry)
XX
XX      Streptavidin gene.
XX
XX      Streptavidin; protein secretion; Bacillus subtilis; ss.
XX
XX      Streptomyces avidinii.
XX
XX      Key      Location/Qualifiers
XX      FH      1..552
XX      FT      CDS
XX      FT      mat_peptide 73..519
XX      FT      sig_peptide 1..72
XX      FT      /tag= b
XX      FT      /tag= c
XX      FT      misc_feature 109..519
XX      FT      /tag= d
XX      FT      /note= "fused to npr signal peptide in pBE673"

WO9324631-A.
XX
XX      09-DEC-1993.
XX
XX      27-MAY-1993; 93WO-US05240.
XX
XX      29-MAY-1992; 92US-0891524.
XX
XX      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX      Nagarajan V;
XX
XX      MPI: 1993-405822/50.
XX      P-PSDB; R44491.
XX
XX      Streptavidin prodn. from Bacillus subtilis - using signal protein
XX      from bacterial exo-protein and expression element from Gram
XX      positive bacterial protein.
XX
XX      Disclosure; Fig 1b; 54pp; English.
```

CC Tetrameric biologically active streptavidin is produced by secretion
CC from *Bacillus subtilis* transformed with a plasmid encoding the
CC sequence.

SQ Sequence 552 BP; 105 A; 207 C; 167 G; 70 T; 3 other;

```
alignment_scores:      Length: 157
      Quality:
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.0000
```

```
alignment_block:
US-09-589-870-2 x Q53412 ..
```

Align seg 1/1 to: Q53412 from: 1 to: 552

27 SerLyAspSerLySAlaInValSerLAlaAGluAGlyIleThrG1 43
79 TCCAAAGACCTCGAAGGCCCAAGGTCCTGGCGCGCCGCGGCAATCCCG 128
43 yThrTPryAsnGlnLeuGlySerThPhelIeValThrAlaGlyAla 60
129 CACCTGGTACAAACAGACTGGCTGCACCTTATCTGTGACCGGGGCCG 178
60 spGlyAlaLeuThrGlyThryThryGlnSerAlaValAGlyAsnAlaGluSer 76
179 ACGGCGCCCTGACCGGAACTCTGAGTGGCGCGTGGCAACGCCAGAGC 228
77 ArgTyValLeuThrGlyArgTyArgTySerAlaProAlaThrAspGlySe 93
93 rGlyThrAlaLeuGlyTrpThValAlaTrpLySAsnAntyArrAsnA 110
279 CGGCACCGCCCTCGTTGGACCGTGGCTGGAAGAATATCTACCGCAACG 328
110 IahIserAlaThrThrTrpSerGlyGlnTyValAGlyAGlyAlaGluAla 126
329 CCCACTCCGGACCAAGTGGACGGCCAGTACGTACGTGGCGCGCGAGGCG 378
127 ArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAlaAsnAl 143
379 AGGATCAACACCCCAAGGGCTGTGACTTCGCCGACACACCGAGGCCAAGC 428
443 aTrpLySerThrLeuValAGlyHisAspThPhenThrLySValLySProS 168
429 CTGGAAAGTCCAGCGTGGTGGCCACAGACACCTTACCAAGGTGAAGCGCT 478
160 eAlaIaAlaSerIleAspAlaAlaLyLySAlaGlyValAsnAsnGlyAsn 176
479 CCGCGCGCTCCATCGACGGCGGAAGAAGCGCGGTCAACACAGGCAAC 528
177 ProLeuAspAlaValAGlnGln 183
329 CCGCTCGACGCGCTTACGACG 549

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q706660

```
seq_documentation_block:
ID   Q70660 standard; DNA; 1296 BP.
```

AC Q70660;

DT 26-MAR-1995 (first entry)

DE SCFV PRAS109 and PRAS113 DNA.

KW Amplification; single chain variable region fusion protein; PCR; ss.

OS	Synthetic.
VY	

EH	Key	Location/qualifiers
0	0	4

FT	CDS	40..1290
FT		/*tag= a
XX		
PN	W09415644-A.	

```
alignment_scores:
  Quality: 140.00
  Ratio: 1.000
  Gaps: 0
Percent Similarity: 100.000
Percent Identity: 100.000
```

alignment_block:
US-09-589-870-2 x Q706660

Align seg 1/1 to: Q70660 from: 1 to: 1296

24 AlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaG1 40
865 GGACAGCCGCTCCAAAGACATCCAAACCTCAGGTTTCTGGAGCGGAAGCTGG 914
40 yIleThrGlyThrTrpTyraSglnLeuGlySerThrPheIleValThra 57
915 TATCACTGGCACCCTGGTATAACCAACTGGGGCGACTTCATTGTGACCG 964
57 laGlyAlaAspGlyAlaLeuThrGlyThrTYRGLuSerAlaValGlyAsn 73
965 CTGGTCGGACGAGAGCTCTGACTGGCACCCTACGAATCTGGCGTGGTATAC 1014
74 AlaGluSerArgTrpValLeuThrGlyThrGTYrAspSerAlaProAlaTh 90
1015 GCAAGATCCCGCTACGACTGACTGCGCGTTATGCTGTGCACCTGGCCAC 1064
90 rAspGlySerGlyThrAlaLeuGlyTYRTPThrValAlaTrpLysAsnAsnT 107
1065 CGATGGCTCTGGACCGCTCTGGGCTGGACGTGACGTGGAAAAACACT 1114
107 yTrArgAsnAlaHisSerAlaThrTrpTrpSerGlyGlnTYrValGlyGly 123
1115 ATCGTATATCGGACACGCGGCACTACGAGTGTGGCGCAATACGTTGGCGGT 1164
124 AlaGluAlaArgGlnAsnThrGlnTrpLeuLeuThrSerGlyThrThnG1 140
1165 GCGAGAGCTCGTATCAACACTAGTGGGCTGTTAACATCCGCGACACTACCGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
1215 AGGAGTGCATGGAAATACACACTAGTAGTCATATACACCTTACCAAG 1264

157 allYpRoSeAlaAlSe 163
|||||
1265 TTAGCCTTCTGCTGCTAGC 1284

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 01:07:06 ; Search time 63.76 Seconds
(without alignments)
194.884 Million cell updates/sec

Title: US-09-589-870-2

Perfect score: 183

Sequence: 1 MRKIVAAIAVSLTTSVITA.....IDAKKAGVNNGNPLDAVQ 183

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 195891 seqs, 67900655 residues

Word size : 0

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_66:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	183	2	A23513 streptavidin precu
2	123	67.2	183	2	S57284 streptavidin V1 pr
3	59	32.2	183	2	S57285 streptavidin V2 pr
4	8	4.4	212	2	T46600 nitrogen metabolic
5	8	4.4	366	2	F81578 DNA polymerase III
6	8	4.4	366	2	H81713 DNA polymerase III
7	8	4.4	366	2	H72090 DNA polymerase III
8	8	4.4	367	2	B83186 probable RND efflu
9	8	4.4	416	2	E71559 probable DNA pol I
10	8	4.4	470	2	D75375 conserved hypotet
11	8	4.4	488	1	S11910 nitrogen metabolic
12	8	4.4	1067	2	G81825 probable drug effl
13	8	4.4	1067	2	E81051 multiple transfe
14	8	4.4	1067	2	E81051 DNA topoisomerase
15	8	4.4	1067	2	S59969 lysin protein S -
16	8	4.4	1067	2	YVBP1 lysin protein S -
17	8	4.4	1067	2	YVBP1 lysin protein S -
18	8	4.4	1067	2	YVBP1 lysin protein S -
19	8	4.4	1067	2	YVBP1 lysin protein S -
20	8	4.4	1067	2	YVBP1 lysin protein S -
21	8	4.4	1067	2	YVBP1 lysin protein S -
22	8	4.4	1067	2	YVBP1 lysin protein S -
23	8	4.4	1067	2	YVBP1 lysin protein S -
24	8	4.4	1067	2	YVBP1 lysin protein S -
25	8	4.4	1067	2	YVBP1 lysin protein S -
26	8	4.4	1067	2	YVBP1 lysin protein S -
27	8	4.4	1067	2	YVBP1 lysin protein S -
28	8	4.4	1067	2	YVBP1 lysin protein S -
29	8	4.4	1067	2	YVBP1 lysin protein S -

30	7	3.8	259	2	D83216 probable aldolase
31	7	3.8	266	1	ELRT1 pancreatic elastas
32	7	3.8	275	2	B49394 translation elonga
33	7	3.8	295	2	T00399 transcription fact
34	7	3.8	300	2	A55792 beta-lactamase (EC
35	7	3.8	319	2	A70605 probable acyl-coa
36	7	3.8	345	2	T08563 dnaJ-related prote
37	7	3.8	376	2	JC6535 multiubiquitin-pro
38	7	3.8	377	2	S63671 26S proteinase cha
39	7	3.8	378	2	S71201 biotin synthase (E
40	7	3.8	397	2	T35609 while protein I - S
41	7	3.8	413	2	F69007 translation elonga
42	7	3.8	433	2	S55968 probable membrane
43	7	3.8	437	2	C72570 probable translati
44	7	3.8	450	1	S13730 pma protein - Esc
45	7	3.8	453	2	T38707 probable initiator

ALIGNMENTS

RESULT 1

A23513 streptavidin precursor - Streptomyces avidinii

C:Species: Streptomyces avidinii

C:Date: 03-Nov-1987 #sequence_1revision 03-Nov-1987 #text_change 10-Dec-1999

C:Accession: A23513; S11540

R:Argarana, C.E.; Kuntz, I.D.; Birken, S.; Axel, R.; Cantor, C.R.

Nucleic Acids Res. 14, 1871-1882, 1986

A:Title: Molecular cloning and nucleotide sequence of the streptavidin gene.

A:Reference number: A23513; MUID:86148514

A:Accession: A23513

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-183 <ARG>

A:Cross-references: GB:X03591; NID:946740; PIDN:CAA27265.1; PID:946741

R:Gittlin, G.; Bayer, E.A.; Mallick, M.

Biochem. J. 269, 527-530, 1990

A:Title: Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine res

A:Reference number: S11540; MUID:90351377

A:Accession: S11540

A:Molecule type: protein

A:Residues: 54-66, 'X', 68 <GIT>

A:Superfamily: streptavidin

C:Keywords: biotin; homotetramer

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-183/Product: streptavidin #status predicted <MAT>

F:67/Binding site: biotin (Tyr) #status predicted

F:78/Binding site: biotin (Tyr) #status predicted

Query Match 100.0%; Score 183; DB 2; Length 183;

Best local Similarity 100.0%; Pred. No. 7.4e-172;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRKIVAAIAVSLTTSVITASADPSKDSKAOVSAEAGITGTWYNQSGFTVAGAD	60
DB	1	MRKIVAAIAVSLTTSVITASADPSKDSKAOVSAEAGITGTWYNQSGFTVAGAD	60
QY	61	GALTGYESAVNAESRYVLTGRYDSAPATDGSCTALGWTAVANKNNYRNAHSATTSQGY	120
DB	61	GALTGYESAVNAESRYVLTGRYDSAPATDGSCTALGWTAVANKNNYRNAHSATTSQGY	120
QY	121	VGAEARINTQWLLTSGTTTANAKSTLVGHDPFTYKPSAASIDAACKAGVNNGNPLDA	180
DB	121	VGAEARINTQWLLTSGTTTANAKSTLVGHDPFTYKPSAASIDAACKAGVNNGNPLDA	180
QY	181	VQO 183	
DB	181	VQO 183	

RESULT 2

S57284
 streptavidin v1 precursor - Streptomyces venezuelae
 C:Species: Streptomyces venezuelae
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999
 C:Accession: S57284
 R:Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.
 Biochim. Biophys. Acta 1263, 60-66, 1995
 A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom
 A:Reference number: S57284; MUID:95359204
 A:Accession: S57284
 A:Molecule type: DNA
 A:Residues: 1-183 <BAY>
 A:Experimental source: strain Tue2460
 C:Superfamily: streptavidin
 F:1-24/Domain: signal sequence #status predicted <STG>
 F:25-183/Product: streptavidin v1 #status predicted <MAT>

Query Match 67.2%; Score 123; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5,3e-113; Indels 0; Gaps 0;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRIYVAIAVSLTVTSITASADPSKSKAQVSAEAGITGTWYNOIGSTFIYTAGAD 60
 - |||||
 Db 1 MKRIYVAIAVSLTVTSITASADPSKSKAQVSAEAGITGTWYNOIGSTFIYTAGAD 60

OY 61 GALTGYESAAGNAESRYVLTGRYDSAPATDGSCTALGWTYAMKNNYRNAHSATTWSGOY 120
 |||||
 Db 61 GALTGYESAAGNAESRYVLTGRYDSAPATDGSCTALGWTYAMKNNYRNAHSATTWSGOY 120

OY 121 VGG 123
 |||
 Db 121 VGG 123

RESULT 3
 S57285
 streptavidin v2 precursor - Streptomyces venezuelae
 C:Species: Streptomyces venezuelae
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999
 C:Accession: S57285
 R:Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.
 Biochim. Biophys. Acta 1263, 60-66, 1995
 A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom
 A:Reference number: S57284; MUID:95359204
 A:Accession: S57285
 A:Molecule type: DNA
 A:Residues: 1-183 <BAY>
 A:Experimental source: strain Tue2605
 C:Superfamily: streptavidin
 F:1-24/Domain: signal sequence #status predicted <STG>
 F:25-183/Product: streptavidin v2 #status predicted <MAT>

Query Match 32.2%; Score 59; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 3,2e-50; Indels 0; Gaps 0;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 63 LTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTYAMKNNYRNAHSATTWSGOY 121
 |||||
 Db 63 LTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTYAMKNNYRNAHSATTWSGOY 121

RESULT 4
 T46600
 nitrogen metabolic regulation protein nmr [imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C>Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 02-Sep-2000
 C:Accession: T46600
 R:Jarral, G.; Marziuf, G.A.
 Curr. Genet. 20, 283-288, 1991
 A:Title: Generation of new mutants of nmr, the negative-acting nitrogen regulatory gene
 A:Reference number: Z23098; MUID:92035098

A:Accession: T46600
 A:Status: preliminary; translated from GB/EMBL/DBDJB
 A:Molecule type: DNA
 A:Residues: 1-212 <JAR>
 A:Cross-references: EMBL:S64286; PIDN:CAB32318.1
 C:Genetics:
 A:Gene: nmr
 C:Superfamily: nitrogen metabolic regulation protein nmr

Query Match 4.4%; Score 8; DB 2; Length 212;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 DAKKAGV 172
 |||||
 Db 83 DAKKAGV 90

RESULT 5
 F81578
 DNA polymerase III, beta chain CP0419 [imported] - Chlamydia pneumoniae (strain A
 C:Species: Chlamydia pneumoniae
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
 C:Accession: F81578
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255
 A:Accession: F81578
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-366 <REA>
 A:Cross-references: GB:AE002203; GB:AE002161; NID:g7189339; PIDN:AAF38262.1; PID:g718
 C:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0419
 C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.4%; Score 8; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVYLTG 82
 |||||
 Db 149 ESRVYLTG 156

RESULT 6
 C81713
 DNA polymerase III, beta chain TC0347 [imported] - Chlamydia muridarum (strain N19g)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
 C:Accession: C81713
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255
 A:Accession: C81713
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-366 <TEF>
 A:Cross-references: GB:AE002302; GB:AE002160; NID:g7190382; PIDN:AAF39208.1; PID:g719
 A:Experimental source: strain N19g (MoPn)
 C:Genetics:
 A:Gene: TC0347
 C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.4%; Score 8; DB 2; Length 366;

Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82
|||||

Db 149 ESRVLTG 156

RESULT 7
H72090
DNA polymerase III (beta chain) - Chlamydomonas reinhardtii (strain CWI029)
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: H72090
R:Kallman, S.; Mitchell, M.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: H72090
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-366 <ARN>
A:Cross-References: GB:AE001618; GB:AE001363; NID:g4376610; PIDN:AA18487.1; PID:g437661
A:Experimental source: strain CWI029
C:Genetics:
A:Gene: dnan
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.4%; Score 8; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82
|||||

Db 149 ESRVLTG 156

RESULT 8
B83186
Probable RND efflux membrane fusion protein precursor PA3677 [Imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: B83186
R:Stover, C.K.; Pham, X.Q.; Ewitt, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
gen.
A:Reference number: A82950
A:Accession: B83186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-367 <STO>
A:Cross-References: GB:AE004787; GB:AE004091; NID:g9949835; PIDN:AA07065.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3677

Query Match 4.4%; Score 8; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 AGVSAAEA 39
|||||

Db 103 AGVSAAEA 110

RESULT 9
E71559
Probable DNA pol III (beta chain) - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 11-Jan-2000
C:Accession: E71559
R:Stephens, R.S.; Kallman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:9900809
A:Accession: E71559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <ARN>
A:Cross-References: GB:AE001282; GB:AE001273; NID:g3328466; PIDN:AA06766.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: dnan
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.4%; Score 8; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82
|||||

Db 199 ESRVLTG 206

RESULT 10
D75375
Conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: D75375
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
., M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: D75375
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-470 <WHI>
A:Cross-References: GB:AE002004; GB:AE000513; NID:g6459366; PIDN:AA11162.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1599
A:Map position: 1

Query Match 4.4%; Score 8; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 57 AGADGALT 64
|||||

Db 392 AGADGALT 399

RESULT 11
S11910
nitrogen metabolic regulation protein nmr - Neurospora crassa
N:Alternate names: nmr protein
C:Species: Neurospora crassa
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S11910; S11924
R:Young, J.L.; Jarai, G.; Fu, Y.H.; Marzluf, G.A.
Mol. Gen. Genet. 222, 120-128, 1990
A:Title: Nucleotide sequence and analysis of NMR, a negative-acting regulatory gene 1
A:Reference number: S11910; MUID:91042412
A:Accession: S11910
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-488 <YOU>

C:Genetics:
A:Gene: nmr
C:Superfamily: nitrogen metabolic regulation protein nmr
C:Keywords: nucleus; transcription regulation

Query Match 4.4%; Score 8; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 DAAKAGV 172
|||||||
DB 181 DAAKAGV 188

RESULT 12
G81825
probable drug efflux protein NMA1969 [Imported] - Neisseria meningitidis (group A strain
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 19-May-2000

C:Accession: G81825

R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
J.; Holtroyd, S.; Jagsis, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556

A:Accession: G81825

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1067 <PAR>

A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85189.1; PID:g738060

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMA1969

C:Superfamily: acriflavin resistance protein

Query Match 4.4%; Score 8; DB 2; Length 1067;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 AAEAGITG 43
|||||||
DB 1054 AAEAGITG 1061

RESULT 13

E81051
multiple transferable resistance system protein MtrD NMB1715 [Imported] - Neisseria men

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000

C:Accession: E81051

R.Rettlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Masiarant, V.; Pizza, M.;
Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Frazer, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: E81051

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1067 <TE>

A:Cross-references: GB:AE002521; GB:AE002098; NID:g7226962; PIDN:AAFA2062.1; PID:g722697

A:Experimental source: serogroup B, strain MMD58

C:Genetics:

A:Gene: NMB1715

C:Superfamily: acriflavin resistance protein

Query Match 4.4%; Score 8; DB 2; Length 1067;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 AAEAGITG 43
|||||||
DB 1054 AAEAGITG 1061

RESULT 14
S59969
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) isoform beta - Chinese hamster

M/Alternate names: DNA topoisomerase II isoform beta; DNA gyrase

C:Species: Cricetus griseus (Chinese hamster)

C>Date: 15-Feb-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999

C:Accession: S59969; S54154

R.Derudder, S.; Frey, S.; Delaporte, C.; Jacquemin-Sablon, A.
Biochim. Biophys. Acta 1264, 178-182, 1995

A:Title: Cloning and characterization of full-length cDNAs coding for the DNA topoisom

A:Reference number: S59969; MUID:96085121

A:Accession: S59969

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-1612 <DER>

A:Cross-references: EMBL:X86455; NID:g790987; PIDN:CAA60173.1; PID:g790988

A:Experimental source: Lung

C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-

C:Keywords: ATP; DNA binding; DNA replication; isomerase; nucleus

F:697-927/Domain: phage T4 DNA topoisomerase (ATP-hydrolyzing) medium chain homology

Query Match 4.4%; Score 8; DB 2; Length 1612;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AGADGALT 64
|||||||
DB 7 AGADGALT 14

RESULT 15

YBPL
Lysis protein S - phage lambda

C:Species: phage lambda

C>Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 23-Jul-1999

C:Accession: H94164; G43012; A04327

R.Daniels, D.

submitted to the Nucleic Acid Sequence Database, September 1982

A:Reference number: A94614

A:Accession: H94164

A:Molecule type: DNA

A:Residues: 1-107 <DNA>

J.Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.
J. Mol. Biol. 162, 729-773, 1982

A:Title: Nucleotide sequence of bacteriophage lambda DNA.

A:Reference number: A92891; MUID:83189071

A:Accession: G43012

A:Molecule type: DNA

A:Residues: 1-107 <SAN>

A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g21510

A>Note: the lambda DNA sequence is from the standard strain lambda-clindts857sam7; t

C:Genetics:

A:Gene: S

A:Map position: 93.16-93.82

C:Superfamily: phage lambda lysis protein S

C:Keywords: host cell lysis

Query Match 3.8%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAKKAGV 172
|||||||
DB 95 AAKKAGV 101

Fri Apr 6 08:41:00 2001

us-09-589-870-2.rpr

Page 5

Search completed: April 6, 2001, 01:26:46
Job time: 1180 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 01:22:46 ; Search time 67.41 Seconds

(Without alignments)
87.670 Million cell updates/sec

Title: US-09-589-870-2
Perfect score: 183
Sequence: 1 MRKIYVAIAVSLTFSITA.....IDAKKAGVNGNPLDAVOQ 183

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 88757 seqs, 32294092 residues

Word size : 0

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	183	1	SAV_STRAV
2	123	67.2	183	1	SAV_STRAV
3	59	32.2	183	1	SAV2_STRVL
4	8	4.4	488	1	NMR_NEUCR
5	8	4.4	1612	1	TP28_CRITO
6	7	3.8	107	1	VLYS_LAMB
7	7	3.8	108	1	VLYL_BPP22
8	7	3.8	132	1	RS8_MICLU
9	7	3.8	151	1	SOD4_MAIZE
10	7	3.8	151	1	SOD5_MAIZE
11	7	3.8	152	1	SODC_CAVPO
12	7	3.8	255	1	CBPM_STRAL
13	7	3.8	266	1	EL1_RAT
14	7	3.8	302	1	DAPA_PROMA
15	7	3.8	376	1	PSD4_MOUSE
16	7	3.8	377	1	PSD4_MOUSE
17	7	3.8	378	1	PSD4_MOUSE
18	7	3.8	379	1	PSD4_MOUSE
19	7	3.8	397	1	WH42_STRCO
20	7	3.8	413	1	EF1A_METTH
21	7	3.8	437	1	EF1A_AERPE
22	7	3.8	450	1	PMB4_ECOLI
23	7	3.8	453	1	YAW6_SCHPO
24	7	3.8	474	1	GSAL_ARATH
25	7	3.8	553	1	DPOL_BPPRD
26	7	3.8	583	1	ARA2_ECOLI
27	7	3.8	637	1	NU5M_STRPU
28	7	3.8	681	1	TKT2_YEAST
29	7	3.8	737	1	CATA_YERPE
30	7	3.8	783	1	OPT3_CAEEL
31	7	3.8	960	1	FGDI_MOUSE
32	7	3.8	1025	1	MK21_YEAST
33	7	3.8	3010	1	POLG_HCVBK

ALIGNMENTS

RESULT ID	SAV_STRAV	STANDARD	PRT	183 AA.
AC	P22629;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	STREPTAVIDIN PRECURSOR.			
OS	Streptomyces avidinii.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomyces; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 24-64.			
RX	MEDLINE-86148514; PubMed-3951999;			
RA	Agarwala C.E., Kuntz J.D., Birken S., Axel R., Cantor C.R.;			
RT	"Molecular cloning and nucleotide sequence of the streptavidin gene.";			
RL	Nucleic Acids Res. 14:1871-1882(1986).			
RN	[2]			
RP	IMPORTANCE OF TRP IN BIOTIN-BINDING.			
RX	MEDLINE-89134083; PubMed-3223904;			
RA	Gittlin G., Bayer E.A., Wilchek M.;			
RT	"Studies on the biotin-binding site of streptavidin. Tryptophan			
RL	residues involved in the active site.";			
RL	Biochem. J. 256:279-282(1988).			
RN	[3]			
RP	IMPORTANCE OF TYR IN BIOTIN-BINDING.			
RX	MEDLINE-90351377; PubMed-2386489;			
RA	Gittlin G., Bayer E.A., Wilchek M.;			
RT	"Studies on the biotin-binding sites of avidin and streptavidin.			
RL	Tyrosine residues are involved in the binding site.";			
RL	Biochem. J. 269:527-530(1990).			
RN	[4]			
RP	CELL-BINDING.			
RX	MEDLINE-90358825; PubMed-2390089;			
RA	Alon R., Bayer E.A., Wilchek M.;			
RT	"Streptavidin contains an RVD sequence which mimics the RGD receptor			
RL	domain of fibronectin.";			
RL	Biochem. Biophys. Res. Commun. 170:1236-1241(1990).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY OF 37-157.			
RX	MEDLINE-89100243; PubMed-2911722;			
RA	Weber P.C., Ohlendorf D.H., Wendoloski J.J., Salameh F.R.;			
RT	"Structural origins of high-affinity biotin binding to streptavidin.";			
RL	Science 243:85-88(1989).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.			
RX	MEDLINE-97337436; PubMed-9194176;			
RA	Freitag S., le Trong I., Klumb L., Stayton P.S., Stenkamp R.E.;			
RT	"Structural studies of the streptavidin binding loop.";			
RL	Protein Sci. 6:1157-1166(1997).			
RN	[7]			
RP	X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS) OF 37-157.			
RX	MEDLINE-97294734; PubMed-9148939;			
RA	Katz B.A., Cass R.T.;			
RT	"In crystals of complexes of streptavidin with peptide ligands			
RT	containing the HPQ sequence the pKa of the peptide histidine is less			

```

RT      than 3.0 "":
RL      J. Biol. Chem. 272:13220-13228(1997).
RN      (8)
RP      X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 37-157.
RX      MEDLINE=98070615; PubMed=9405158;
RA      Katz B.A.;
RT      "Binding of biotin to streptavidin stabilizes intersubunit salt
RT      bridges between Asp61 and His67 at low pH.";
RL      J. Mol. Biol. 274:776-800(1997).
RN      (9)
RP      X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 37-157 OF MUTANTS.
RX      MEDLINE=98300340; PubMed=9636711;
RA      Freitag S., le Trong I., Chikoti A., Klumb L.A., Stayton P.S.,
RT      Stenkamp R.E.;
RT      "Structural studies of binding site tryptophan mutants in the high-
RT      affinity streptavidin-biotin complex.";
RL      J. Mol. Biol. 279:211-221(1998).
CC      -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC      -1- FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC      MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC      -1- SUBUNIT: HOMOTETRAMER.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC      -1- DATABASE: NAME=Prozyme technical fact sheet;
CC      WWW="http://www.prozyme.com/technical/salodata.html".
CC      -----
DR      EMBL; X03591; CAA27265.1; -.
DR      PIR; A23513; A23513.
DR      PDB; 1STP; 15-OCT-94.
DR      PDB; 1PTS; 15-OCT-94.
DR      PDB; 1SRE; 01-NOV-94.
DR      PDB; 1SRF; 01-NOV-94.
DR      PDB; 1SRG; 01-NOV-94.
DR      PDB; 1SRH; 01-NOV-94.
DR      PDB; 1SRI; 01-NOV-94.
DR      PDB; 1SRJ; 01-NOV-94.
DR      PDB; 1RSU; 15-MAY-97.
DR      PDB; 1SLD; 03-APR-96.
DR      PDB; 1SLE; 03-APR-96.
DR      PDB; 1SLP; 03-APR-96.
DR      PDB; 1SLG; 03-APR-96.
DR      PDB; 1STR; 08-MAR-96.
DR      PDB; 1STS; 08-MAR-96.
DR      PDB; 1SMA; 04-MAR-98.
DR      PDB; 1SMB; 04-MAR-98.
DR      PDB; 1SMC; 04-MAR-98.
DR      PDB; 1SMD; 04-MAR-98.
DR      PDB; 1SWE; 04-MAR-98.
DR      PDB; 1SWG; 15-JUL-98.
DR      PDB; 1SMH; 23-MAR-99.
DR      PDB; 1SMJ; 23-MAR-99.
DR      PDB; 1SMK; 23-MAR-99.
DR      PDB; 1SMN; 23-MAR-99.
DR      PDB; 1SMO; 23-MAR-99.
DR      PDB; 1SWP; 23-MAR-99.
DR      PDB; 1SMQ; 23-MAR-99.
DR      PDB; 1SMR; 23-MAR-99.
DR      PDB; 1VWA; 18-MAR-98.
DR      PDB; 1VWB; 18-MAR-98.
DR      PDB; 1VWC; 18-MAR-98.
DR      PDB; 1VWD; 18-MAR-98.
DR      PDB; 1VWE; 18-MAR-98.
DR      PDB; 1VWF; 18-MAR-98.

```

DR	PDB; 1VWG; 18-MAR-98.
DR	PDB; 1VWH; 18-MAR-98.
DR	PDB; 1VWI; 18-MAR-98.
DR	PDB; 1VWJ; 18-MAR-98.
DR	PDB; 1VWK; 18-MAR-98.
DR	PDB; 1VWL; 18-MAR-98.
DR	PDB; 1VWM; 18-MAR-98.
DR	PDB; 1VWN; 18-MAR-98.
DR	PDB; 1VWO; 18-MAR-98.
DR	PDB; 1VWP; 18-MAR-98.
DR	PDB; 1VWR; 18-MAR-98.
DR	PDB; 1SWF; 29-APR-98.
DR	PDB; 21ZA; 16-SEP-98.
DR	PDB; 21ZB; 16-SEP-98.
DR	PDB; 21ZC; 16-SEP-98.
DR	PDB; 21ZD; 23-SEP-98.
DR	PDB; 21ZE; 16-SEP-98.
DR	PDB; 21ZF; 16-SEP-98.
DR	PDB; 21ZG; 16-SEP-98.
DR	PDB; 21ZH; 16-SEP-98.
DR	PDB; 21ZI; 16-SEP-98.
DR	PDB; 21ZJ; 16-SEP-98.
DR	PDB; 21ZK; 16-SEP-98.
DR	PDB; 21ZL; 16-SEP-98.
DR	PDB; 2RTA; 18-NOV-98.
DR	PDB; 2RTB; 18-NOV-98.
DR	PDB; 2RTC; 18-NOV-98.
DR	PDB; 2RTD; 18-NOV-98.
DR	PDB; 2RTE; 18-NOV-98.
DR	PDB; 2RTF; 27-JAN-99.
DR	PDB; 2RTG; 18-NOV-98.
DR	PDB; 2RTH; 18-NOV-98.
DR	PDB; 2RTI; 25-NOV-98.
DR	PDB; 2RTJ; 25-NOV-98.
DR	PDB; 2RTK; 18-NOV-98.
DR	PDB; 2RTL; 18-NOV-98.
DR	PDB; 2RTM; 18-NOV-98.
DR	PDB; 2RTN; 18-NOV-98.
DR	PDB; 2RTO; 18-NOV-98.
DR	PDB; 2RTP; 18-NOV-98.
DR	PDB; 2RTO; 18-NOV-98.
DR	PDB; 2RTR; 18-NOV-98.
DR	INTERPRO; IPR000088; .
DR	PFAM; PF01382; Avidin; 1.
DR	PROSITE; PS00577; AVIDIN; 1.
KW	Signal; Biotin; 3d-structure.
FT	SIGNAL 1 24
FT	CHAIN 25 183 STREPTAVIDIN.
FT	BINDING 67 67 INVOLVED IN BIOTIN BINDING.
FT	BINDING 78 78 INVOLVED IN BIOTIN BINDING.
FT	BINDING 116 116 INVOLVED IN BIOTIN BINDING.
FT	BINDING 132 132 INVOLVED IN BIOTIN BINDING.
FT	BINDING 144 144 INVOLVED IN BIOTIN BINDING.
FT	SITE 83 85 CELL ATTACHMENT SITE.
FT	HELIX 38 41
FT	STRAND 43 47
FT	TURN 48 49
FT	STRAND 52 57
FT	TORN 59 60
FT	STRAND 62 68
FT	STRAND 77 84
FT	STRAND 95 104
FT	STRAND 109 121
FT	STRAND 127 136
FT	HELIIX 140 145
FT	STRAND 147 156
SEQ	SEQUENCE 183 AA; 18834 MW; 4652D8AE018468F5 CRC64;

Query Match 100.0%; Score 183; DB 1; length 183;
 Best Local Similarity 100.0%; Pred. No. 1,5e-17;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MRRIVAAIVSLTTVSITTSASADPSKDSKAOVSAEAGITGTWYNQLGSTFIVTAGAD 60
Db 1 MRRIVAAIVSLTTVSITTSASADPSKDSKAOVSAEAGITGTWYNQLGSTFIVTAGAD 60
OY 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTAVAMKNYRNHSAATTMSGOY 120
Db 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTAVAMKNYRNHSAATTMSGOY 120
OY 121 VGGAERINQWLLTSTTANNAKSTLVGHDPFTTKPSASATDAKKGAVNNGNRLDA 180
Db 121 VGGAERINQWLLTSTTANNAKSTLVGHDPFTTKPSASATDAKKGAVNNGNRLDA 180
OY 181 VQG 183
Db 181 VQG 183

RESULT 2
SAV2_STRVL STANDARD: PRT: 183 AA.
ID SAV1_STRVL
AC 053532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
OS STREPTAVIDIN V1 PRECURSOR (SA V1).
OC Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95359204; Pubmed=7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RL from Streptomyces."
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S78777; AAB35015.1; -.
CC HSSP: P22629; IPTS.
CC INTERPRO: IPR000088; -.
CC PFM: PF01382; Avidin; 1.
CC PROSITE: PS00577; AVIDIN; 1.
CC KW Signal; Biotin.
CC FT SIGNAL 1 24 BY SIMILARITY.
CC FT CHAIN 25 183 STREPTAVIDIN V1.
CC FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 78 78 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC SQ SEQUENCE 183 AA: 18864 MW: 4652088B14847DE0 CRC64;

Query Match 67.2%; Score 123; DB 1; Length 183;
```

```

Best Local Similarity 100.0%; Pred. No. 6.4e-117;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRRIVAAIVSLTTVSITTSASADPSKDSKAOVSAEAGITGTWYNQLGSTFIVTAGAD 60
Db 1 MRRIVAAIVSLTTVSITTSASADPSKDSKAOVSAEAGITGTWYNQLGSTFIVTAGAD 60
OY 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTAVAMKNYRNHSAATTMSGOY 120
Db 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTAVAMKNYRNHSAATTMSGOY 120
OY 121 VQG 123
Db 121 VQG 123

RESULT 3
SAV2_STRVL STANDARD: PRT: 183 AA.
ID SAV1_STRVL
AC 053533;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE STREPTAVIDIN V2 PRECURSOR (SA V2).
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95359204; Pubmed=7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RL from Streptomyces."
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S78782; AAB35016.1; -.
CC HSSP: P22629; IPTS.
CC INTERPRO: IPR000088; -.
CC PFM: PF01382; Avidin; 1.
CC PROSITE: PS00577; AVIDIN; 1.
CC KW Signal; Biotin.
CC FT SIGNAL 1 24 BY SIMILARITY.
CC FT CHAIN 25 183 STREPTAVIDIN V2.
CC FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 78 78 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC SQ SEQUENCE 183 AA: 18833 MW: FEARFFDEA4ECCA CRC64;

Query Match 32.2%; Score 59; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 3.1e-52;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

OY 63 LTGTESAVGNABESRYVLGRYDSAPATDGGTALGWTVMKNNYRNAHSATWSGOYV 121
    |||||||
DB 63 LTGTESAVGNABESRYVLGRYDSAPATDGGTALGWTVMKNNYRNAHSATWSGOYV 121

RESULT 4
NMR_NEUCR
ID NMR_NEUCR STANDARD: PRT: 488 AA.
AC P23762;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NITROGEN METABOLIC REGULATION PROTEIN (NMR PROTEIN).
GN NMR.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=91042412; PubMed=2146484;
RA Young J.L., Jaral G., Fu Y.H., Marzluf G.A.;
RT "Nucleotide sequence and analysis of NMR, a negative-acting
RL regulatory gene in the nitrogen circuit of Neurospora crassa.";
RM Mol. Genet. 222:120-128(1990).
RN [2]
RP SEQUENCE OF 99-310 FROM N.A.
RX MEDLINE=92035098; PubMed=1834354;
RA Jaral G., Marzluf G.A.;
RT "Generation of new mutants of nmr, the negative-acting nitrogen
RL regulatory gene of Neurospora crassa, by repeat induced mutation.";
RM Curr. Genet. 20:283-288(1991).
RN [3]
RP MUTAGENESIS.
RC STRAIN=74-OR23-1A;
RX MEDLINE=91109706; PubMed=2148799;
RA Jaral G., Marzluf G.A.;
RT "Analysis of conventional and in vitro generated mutants of nmr, the
RL negatively acting nitrogen regulatory gene of Neurospora crassa.";
RM Mol. Genet. 222:233-240(1990).
CC -I- FUNCTION: MAJOR NEGATIVE REGULATORY PROTEIN IN THE NITROGEN
CC CONTROL CIRCUIT OF NEUROSPORA CRASSA WHICH ALLOWS UTILIZATION
CC OF VARIOUS NITROGEN-CONTAINING COMPOUNDS.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S64286; CAB32318.1; -.
CC DR PIR: S11910; S11910.
CC DR Transcription regulation; Repressor; Nitrate assimilation;
CC KW Nuclear protein.
CC FT MUTAGEN 2 2 P->K: LOSS OF FUNCTION.
CC FT MUTAGEN 386 386 G->D,R: LOSS OF FUNCTION.
CC FT MUTAGEN 386 386 G->C,S: FEW, VERY SLOWLY GROWING
CC FT TRANSFORMANTS.
CC FT SIMILAR 96 193 TO YEAST ARGRTII (AA 588-681).
CC FT DOMAIN 412 488 DISPENSABLE FOR NMR FUNCTION.
CC FT SEQUENCE 488 AA; 54857 MW; DDDAF91671C7846C CRC64;

Query Match 4.4%; Score 8; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 165 DAAKAGV 172
    |||||||
DB 161 DAAKAGV 188

```

```

RESULT 5
TP2B_CRILLO
ID TP2B_CRILLO STANDARD: PRT: 1612 AA.
AC Q64399;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN TOP2B.
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=96085121; PubMed=7495861;
RA Derudder S., Frey S., Delaporte C., Jacquemin-sablon A.;
RT "Cloning and characterization of full-length cDNAs coding for the DNA
RT topoisomerase II beta from Chinese hamster lung cells sensitive and
RL resistant 9-OH-ellipticine.";
RM Biochim. Biophys. Acta 1264:178-182(1995).
CC -I- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -I- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -I- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X86455; CA60173.1; -.
CC DR HSSP: P06786; IBGW.
CC DR INTERPRO: IPR000947; -.
CC DR INTERPRO: IPR001154; -.
CC DR INTERPRO: IPR001241; -.
CC DR INTERPRO: IPR002205; -.
CC DR PFAM: PF00204; DNA_topoisoi1; 1.
CC DR PFAM: PF00521; DNA_topoisoi4; 1.
CC DR PRINTS: PR00418; TP12FAMILY.
CC DR PRINTS: PR00615; CCAATSUBUNTA.
CC DR PRINTS: PR01158; TOPISMRASEII.
CC DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
CC KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
CC FT NP_BIND 170 175 ATP (POTENTIAL).
CC FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
CC FT SEQUENCE 1612 AA; 182074 MW; C01D6FC40620FC68 CRC64;

Query Match 4.4%; Score 8; DB 1; Length 1612;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 57 AGADGALT 64
    |||||||
DB 7 AGADGALT 14

RESULT 6
VLXS_LAMB

```



```

ID VLVS_LAMB     STANDARD:      PRT:      107 AA.
AC P03705;
DT 21-JUL-1966 (Rel. 01, Created)
DT 21-JUL-1966 (Rel. 01, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOLIN (GPS PROTEIN) [INCLUDES: LYSIS PROTEIN S; LYSIS INHIBITOR].
GN S.
OS Bacteriophage lambda.
OC Viruses: dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA."
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=86304148; PubMed=2943725;
RA Raab R., Neal G., Garrett J., Grimalia R., Fusselman R., Young R.;
RT "Mutational analysis of bacteriophage lambda lysis gene S."
RL J. Bacteriol. 167:1035-1042(1986).
CC -1- FUNCTION: THE 105 AA PRODUCT INDUCES HOST CELL LYSIS BY CAUSING
CC FORMATION OF PORES IN THE INNER MEMBRANE. THE 107 AA PROTEIN
CC COUNTERACTS THE FORMATION OF SUCH TUNNELS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, INNER MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: USING ALTERNATIVE INITIATION CODONS IN
CC THE SAME READING FRAME, THE GENE TRANSLATES INTO TWO PRODUCTS:
CC THE LYSIS PROTEIN S AND AN INHIBITOR OF LYSIS.
CC -1- SIMILARITY: BELONGS TO THE LAMBDA PHAGE S PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J02459; AAA96597.1; -.
DR EMBL: M14035; AAA32248.1; -.
DR PIR: A04327; YVBPL.
KW Phage lysis protein; Transmembrane; Inner membrane;
KM Alternative Initiation.
FT CHAIN 1 107 LYSIS INHIBITOR.
FT INIT_MET 3 3 FOR LYSIS PROTEIN S.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
SQ SEQUENCE 107 AA; 11520 MW; 66D0D62426F1766E CRC64;

Query Match 3.8%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 AAKKAGV 172
DB 95 AAKKAGV 101

RESULT 7
VLT1_BPP22 STANDARD: PRT: 108 AA.
ID VLT1_BPP22
AC P09962;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOLIN (LYSIS PROTEIN 13).
GN 13.
OS Bacteriophage P22.
OC Viruses: dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.

```

```

RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=86045883; PubMed=2998005;
RA Remell D., Potete A.R.;
RT "Phage P22 lysis genes: nucleotide sequences and functional
RT relationships with T4 and lambda genes."
RL Virology 143:280-289(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Schickmaler P., Huber J., Schmeleger H.;
RL Submitted (Dec-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX Kropinski A.M.B., VanderByl C.S.;
RT "The completed sequence of genome of Salmonella phage P22."
RL Submitted (Dec-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR LYSIS OF THE BACTERIAL CELL WALL BY
CC DISRUPTING THE CELL MEMBRANE, THEREBY GIVING HYDROLYTIC ENZYMES
CC ACCESS TO THE CELL WALL.
CC -1- SIMILARITY: BELONGS TO THE LAMBDA PHAGE S PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M10997; AAA23265.1; -.
DR EMBL: X67137; CAA47616.1; -.
DR EMBL: AF217253; AAF75039.1; -.
DR PIR: S22903; YVBPS2.
KW Phage lysis protein; Late protein.
SQ SEQUENCE 108 AA; 11682 MW; 6233F931AE7C787E CRC64;

Query Match 3.8%; Score 7; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 AAKKAGV 172
DB 96 AAKKAGV 102

RESULT 8
RS8_MICLU STANDARD: PRT: 132 AA.
ID RS8_MICLU
AC P33106;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S8.
GN RPSH.
OS Micrococcus luteus (Micrococcus lysodeikticus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Micrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90133967; PubMed=2533272;
RA Ohama T., Muto A., Osawa S.;
RT "Spectinomycin operon of Micrococcus luteus: evolutionary
RT implications of organization and novel codon usage."
RL J. Mol. Evol. 29:381-395(1989).
CC -1- FUNCTION: BINDS DIRECTLY TO THE CENTRAL DOMAIN OF 16S RIBOSOMAL
CC RNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X17524; CAA35561.1; -.

DR PIR: S29885; S29885.

DR HSSP: P56209; 1SET.

DR INTERPRO: IPR000630; -.

DR PFAM: PF00410; Ribosomal_S8; 1.

DR PROSITE: PS00053; RIBOSOMAL_S8; 1.

DR Ribosomal protein; rRNA-binding.

KW INIT_MET 0

FT BY SIMILARITY.

SQ SEQUENCE 132 AA; 14363 MW; 0EDB952F12C3143D CRC64;

Query Match 3.8%; Score 7; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 AAKAGV 172
Db 117 AAKAGV 123

RESULT 9
SOD4_MAIZE STANDARD; PRT; 151 AA.

AC P23345;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SUPEROXIDE DISMUTASE [CU-ZN] 4A (EC 1.15.1.1).
GN SODCC.3 OR SOD4P.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP MEDLINE:90136495; PubMed:2482436;
RA Cannon R.E., Scandalios J.G.;
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
CC -!- SUBUNIT: HOMODIMER.
CC -!- CELLULAR LOCATION: CYTOPLASMIC.
CC -!- MISCELLANEOUS: EUKARYOTIC CELLS CONTAIN A MITOCHONDRIAL MN-CONTAINING ENZYME & A CYTOPLASMIC CU-ZN-CONTAINING ENZYME.
CC PLANT ALSO HAVE A CHLOROPLAST CU-ZN ENZYME.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X17564; -; NOT_ANNOTATED_CDS.
DR PIR: S07007; S07007.
DR HSSP: P07505; 1SRD.
DR MAIZEDB: 47586; -.
DR PFAM: PF00080; SODCC.3.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR PROSITE: PS00087; SOD_CU_ZN_1; 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW Oxidoreductase; Copper; Zinc; Multigene family.
FT INIT_MET 0

FT METAL 44 44 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT DISULFID 55 144 BY SIMILARITY.
SQ SEQUENCE 151 AA; 14983 MW; 9C6226F86C919E58 CRC64;

Query Match 3.8%; Score 7; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 VTAGAG 61
Db 85 VTAGAG 91

RESULT 10
SOD5_MAIZE STANDARD; PRT; 151 AA.

AC P23346;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SUPEROXIDE DISMUTASE [CU-ZN] 4AP (EC 1.15.1.1).
GN SODCC.2 OR SOD4AP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP MEDLINE:90136495; PubMed:2482436;
RA Cannon R.E., Scandalios J.G.;
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
CC -!- SUBUNIT: HOMODIMER.
CC -!- CELLULAR LOCATION: CYTOPLASMIC.
CC -!- MISCELLANEOUS: EUKARYOTIC CELLS CONTAIN A MITOCHONDRIAL MN-CONTAINING ENZYME & A CYTOPLASMIC CU-ZN-CONTAINING ENZYME.
CC PLANT ALSO HAVE A CHLOROPLAST CU-ZN ENZYME.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X17565; -; NOT_ANNOTATED_CDS.
DR PIR: S07008; S07008.
DR HSSP: P07505; 1SRD.
DR MAIZEDB: 47586; -.
DR MENDEL: 298; ZEAMA; SODCC.2.
DR INTERPRO: IPR001424; -.
DR PFAM: PF00080; SODCC.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR PROSITE: PS00087; SOD_CU_ZN_1; 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW Oxidoreductase; Copper; Zinc; Multigene family.
FT INIT_MET 0
FT METAL 44 44 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).

FT METAL 78 78 ZINC (BY SIMILARITY).
 FT METAL 81 81 ZINC (BY SIMILARITY).
 FT METAL 118 118 COPPER (BY SIMILARITY).
 FT DISULFID 55 144 BY SIMILARITY.
 SQ SEQUENCE 151 AA; 14939 MW; 9C7E572A6C1AEF1D CRC64;

Query Match 3.8%; Score 7; DB 1; Length 151;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 VTGADG 61
 |||||
 Db 85 VTGADG 91

RESULT 11
 SODC_CAVPO STANDARD; PRT; 152 AA.
 AC P33431;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE SUPEROXIDE DISMUTASE [CU-ZN] (EC 1.15.1.1).
 GN SOD1.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 RN Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 RP SEQUENCE FROM N.A.
 RC STRAIN-HARTLEY; TISSUE-LUNG;
 RA Yuan H.T., Bingle C.D., Kelly F.J.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE:
 RC TISSUE-LIVER;
 MEDLINE=94059414; PubMed=8240718;
 RA Wolf B., Reinecke K., Aumann K.-D., Brigelius-Flohe R., Flohe L.;
 RT "Taxonomical classification of the guinea pig based on its Cu/Zn
 superoxide dismutase sequence.";
 RL Biol. Chem. Hoppe-Seyler 374:641-649(1993).
 CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
 CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- MISCELLANEOUS: EUKARYOTIC CELLS CONTAIN A MITOCHONDRIAL
 MN-CONTAINING ENZYME & A CYTOPLASMIC CU-ZN-CONTAINING ENZYME.
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U39844; AAC52720.1; -.
 DR PIR: S36108.
 DR HSP: P00441; 1SPD.
 DR INTERPRO: IPR001424; -.
 DR PFAM: PF00080; sodbu.1.
 DR PRINTS: PRO0068; CUZNDISMUTASE.
 DR PROSITE: PS00087; SOD_CU_ZN_1; 1.
 DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
 KW Oxidoreductase; Copper; Zinc; Acetylation.
 FT INIT_MER 0 0
 FT MOD_RES 1 1 ACETYLATION.
 FT METAL 45 45 COPPER (BY SIMILARITY).
 FT METAL 47 47 COPPER (BY SIMILARITY).
 FT METAL 62 62 COPPER AND ZINC (BY SIMILARITY).
 FT METAL 70 70 ZINC (BY SIMILARITY).

FT METAL 79 79 ZINC (BY SIMILARITY).
 FT METAL 82 82 ZINC (BY SIMILARITY).
 FT METAL 119 119 COPPER (BY SIMILARITY).
 FT DISULFID 56 145 BY SIMILARITY.
 FT CONFLICT 102 103 LI -> IL (IN REF. 2).
 SQ SEQUENCE 152 AA; 15538 MW; 5891CE16E2ACBFB3 CRC64;

Query Match 3.8%; Score 7; DB 1; Length 152;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 VTGADG 61
 |||||
 Db 86 VTGADG 92

RESULT 12
 CBPM_STRAL STANDARD; PRT; 255 AA.
 AC P00733;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MURAMOYL-PENTAPEPTIDE CARBOXYPEPTIDASE PRECURSOR (EC 3.4.17.8)
 DE (METALLO DD-PEPTIDASE) (ZN DD-PEPTIDASE) (D-ALANYL-D-ALANINE
 DE CARBOXYPEPTIDASE).
 OS Streptomyces albus G.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 RN Streptomycetaceae; Streptomyces.
 RP SEQUENCE FROM N.A.
 RC STRAIN-G;
 RA MEDLINE=9115060; PubMed=2276609;
 RA Duez C., Lakaye B., Houba S., Dusart J., Ghuyssen J.-M.;
 RT "Cloning, nucleotide sequence and amplified expression of the gene
 RT encoding the extracellular metallo (Zn) DD-peptidase of Streptomyces
 RT albus G.";
 RL FEBS Microbiol. Lett. 59:215-219(1990).
 RN [2]
 RP SEQUENCE OF 43-255.
 RC STRAIN-SOLVIFACIENS;
 RX MEDLINE=83131648; PubMed=6825689;
 RA Joris B., van Beumen J., Casagrande F., Gerday C., Frere J.-M.,
 RA Ghuyssen J.-M.;
 RT "The complete amino acid sequence of the Zn2+-containing D-alanyl-D-
 RT alanine-cleaving carboxypeptidase of streptomyces albus G.";
 RL Eur. J. Biochem. 130:53-69(1983).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=83012968; PubMed=7121586;
 RA Dideberg O., Charlier P., Dive G., Joris B., Frere J.-M.,
 RA Ghuyssen J.-M.;
 RT "Structure of a Zn2+-containing D-alanyl-D-alanine-cleaving
 RT carboxypeptidase at 2.5-A resolution.";
 RL Nature 299:469-470(1982).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RA Wery J.-P., Charlier P., Dideberg O.;
 RL Submitted (MAR-1996) to the PDB data bank.
 CC -1- FUNCTION: THIS ENZYME CATALYZES CARBOXYPEPTIDATION AND
 TRANSEPTIDATION REACTIONS INVOLVED IN BACTERIAL CELL WALL
 METABOLISM. IT EFFECTIVELY CATALYZES THE TRANSFER OF THE N-ALPHA,
 N-EPSILON-DIACETYL-L-LYSYL-D-ALANYL ELECTROPHILIC GROUP OF THE
 CC STANDARD TRIPEPTIDE SUBSTRATE N-ALPHA,N-EPSILON-DIACETYL-L-LYSYL-
 CC D-ALANYL-D-ALANINE TO WATER. IT ALSO PERFORMS A WEAK BETA-
 CC LACTAMASE ACTIVITY, HYDROLYSING PENICILLIN INTO PENICILLATE AT A
 CC VERY LOW RATE.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES THE C-TERMINAL D-ALANYL-D-ALANINE
 CC BOND IN PEPTIDOGLYCAN.
 CC -1- COFACTOR: ZINC.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- PTM: THE N-TERMINUS IS PARTIALLY BLOCKED AS A RESULT OF THE

```

CC CYCLIZATION OF THE FIRST TWO AA INTO ANHYDROSPARTYLGLYCINE IMIDE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X55794; CAA39319.1; -.
CC PIR: A00913; CPSPMU.
CC PDB: 1LBV; 08-NOV-96.
CC MEROPS: M15.001; -.
CC INTERPRO: IPR002477; -.
CC PFAM: PF01471; PG_binding_1; 1.
CC Hydrolase; Carboxypeptidase; Cell wall; Zinc; 3D-structure; Signal.
CC SIGNAL 1 42
CC CHAIN 43 255
CC MOD_RES 43 43
CC DISULFID 45 123
CC DISULFID 136 184
CC DISULFID 212 253
CC METAL 196 196
CC METAL 237 237
CC METAL 239 239
CC BINDING 180 180
CC ACT_SITE 234 234
CC CONFLICT 43 43
CC CONFLICT 110 109
CC SIMILAR 43 109
CC SO SEQUENCE 255 AA; 26190 MW; 9168C2145A863AD3 CRC64;

```

Query Match
Best Local Similarity 3.8%; Score 7; DB 1; Length 255;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 19 TASASAD 25
Db 32 TASASAD 38

RESULT 13
EIL_RAT
ID EIL_RAT STANDARD: PRT; 266 AA.
AC P00773;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).
GN ELA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Pictet R.L.,
RA Nikovits W., Rutter W.J.;
RT "Primary structure of two distinct rat pancreatic preproelastases
RT determined by sequence analysis of the complete cloned messenger
RT ribonucleic acid sequences";
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Iahae R.G.,
RA Rutter W.J., MacDonald R.J.;
RT "Structure of the two related elastase genes expressed in the rat
RT pancreas";
RL J. Biol. Chem. 259:14271-14278(1984).

```

RN [3]
RN SEQUENCE OF 17-45.
RP TISSUE=PANCREAS;
RX MEDLINE=84000385; PubMed=6555050;
RA Larrman C.;
RT Isolation and characterization of rat pancreatic elastase.";
RL Biochemistry 22:3763-3770(1983).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.
CC -1- SUBCELLULAR LOCATION: ALA-1-XAA.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V01234; CAA24544.1; -.
CC EMBL: L00117; AAA98811.1; -.
CC EMBL: L00112; AAA98811.1; JOINED.
CC EMBL: L00113; AAA98811.1; JOINED.
CC EMBL: L00114; AAA98811.1; JOINED.
CC EMBL: L00115; AAA98811.1; JOINED.
CC EMBL: L00116; AAA98811.1; JOINED.
CC PIR: A00960; ELRT1.
CC HSSP: P00772; 1EST.
CC MEROPS: S01.153; -.
CC INTERPRO: IPR001254; -.
CC INTERPRO: IPR001314; -.
CC PFAM: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; Pancreas; Zymogen; Signal.
CC SIGNAL 1 16
CC PROPEP 17 26
CC CHAIN 27 266
CC DISULFID 56 72
CC DISULFID 153 220
CC DISULFID 184 200
CC DISULFID 210 240
CC ACT_SITE 71 71
CC ACT_SITE 119 119
CC ACT_SITE 214 214
CC CONFLICT 104 104
CC CONFLICT 108 108
CC CONFLICT 244 244
CC CONFLICT 266 266
CC SO SEQUENCE 266 AA; 28976 MW; 5A56F8FCF1AEDA CRC64;

```

Query Match
Best Local Similarity 3.8%; Score 7; DB 1; Length 266;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 121 VGGAER 127
Db 28 VGGAER 34

RESULT 14
DAPA_PROMA
ID DAPA_PROMA STANDARD: PRT; 302 AA.
AC P49423;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)

```

DE DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52) (DHDPs).
GN DAPA.
OS Prochlorococcus marinus.
OC Bacteria: Cyanobacteria; Prochlorophytes; Prochlorococaceae;
OC Prochlorococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCMP 1375;
RA Lorenz M., Boerner T., Hess W.R.;
RT "Molecular cloning and characterization of a dihydrodipicolinate
RT synthase (DHDPs) gene from the photoautotrophic prokaryote
RT Prochlorococcus marinus CCMP 1375 (Prochlorophyta).";
RL Endocyt. Cell Res. 11:59-68(1995).
CC -1- CATALYTIC ACTIVITY: L-ASPARTATE 4-SEMIALDEHYDE + PYRUVATE =
CC DIHYDRODIPICOLINATE + 2 H(2)O.
CC -1- ENZYME REGULATION: SENSITIVE TO LYSINE INHIBITION
CC (BY SIMILARITY).
CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DHDPs FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: Z68126; CAA92211.1; -.
CC DR HSP: P05640; LDHP.
CC DR INTERPRO: IPR001899; -.
CC DR INTERPRO: IPR002220; -.
CC DR PFAM: PF00701; DHDPs_1.
CC DR PRINTS: PR00146; DHPICSNTHASE.
CC DR PROSITE: PS00665; DHDPs_1; 1.
CC DR PROSITE: PS00666; DHDPs_2; 1.
CC KM Lyase; Diaminopimelate biosynthesis; lysine biosynthesis;
CC Feedback-inhibition.
CC FT ACT_SITE 172 172 BY SIMILARITY.
CC SQ SEQUENCE 302 AA; 31769 MW; F11F6023643FB84 CRC64;

Query Match 3.8%; Score 7; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AGADGAL 63
DB 105 AGADGAL 111

RESULT 15
PSD4_MOUSE
ID PSD4_MOUSE STANDARD; PRT; 376 AA.
AC 035226;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 26S PROTEASOME REGULATORY SUBUNIT 55A (MULTIUBQUITIN CHAIN BINDING
DE PROTEIN).
GN PSMD4 OR MCB1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98112730; PubMed-9511739;
RA Pusch W., Jaehner D., Ivell R.;
RT "Molecular cloning and testicular expression of the gene transcripts
RT encoding the murine multiubiquitin-chain-binding protein (Mcb1).";

```

```

RL Gene 207:19-24(1998).
CC -1- FUNCTION: BINDS AND PRESUMABLY SELECTS UBIQUITIN-CONJUGATES FOR
CC DESTRUCTION.
CC -1- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT 55A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC DR EMBL: AF013099; AAC53547.1; -.
CC DR MGD; MGI:1201670; PSMD4.
CC KW Proteasome.
CC SQ SEQUENCE 376 AA; 40703 MW; 732AC02B56760EAA CRC64;

```

```

Query Match 3.8%; Score 7; DB 1; Length 376;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SAAEAGI 41
DB 242 SAAEAGI 248

```

Search completed: April 6, 2001, 01:29:25
 Job time: 399 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 6, 2001, 01:12:46 : Search time 67.82 Seconds
(without alignments)
316.264 Million cell updates/sec

Title: US-09-589-870-2
Perfect score: 183
Sequence: 1 MRKIVAAIAVSLFTVSITA.....IDAKKAGVNGNPLDAVQ 183

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size : 0 374700

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_15:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp_invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	4.4	271	10 Q9M8S8	Q9M8S8 arabidopsis
2	8	4.4	366	2 Q9Z8K0	Q9Z8K0 chlamydia p
3	8	4.4	366	2 Q9PKW4	Q9PKW4 chlamydia m
4	8	4.4	366	2 Q9K274	Q9K274 chlamydia p
5	8	4.4	366	2 Q9JSG3	Q9JSG3 chlamydia p
6	8	4.4	416	2 Q84078	Q84078 chlamydia t
7	8	4.4	428	3 Q99006	Q99006 trichoderma
8	8	4.4	470	2 Q9RTZ8	Q9RTZ8 deinococcus
9	8	4.4	1067	2 Q9JY67	Q9JY67 neisseria m
10	8	4.4	1067	2 Q9JY51	Q9JY51 neisseria m
11	7	3.8	86	2 Q9WTF8	Q9WTF8 xanthomonas
12	7	3.8	86	2 Q9RBK3	Q9RBK3 xanthomonas
13	7	3.8	86	2 Q9K363	Q9K363 xanthomonas
14	7	3.8	106	2 Q9MGN5	Q9MGN5 bacterioph
15	7	3.8	107	9 Q9MCT3	Q9MCT3 bacterioph
16	7	3.8	117	10 Q9SL91	Q9SL91 arabidopsis
17	7	3.8	124	10 P93800	P93800 zea mays (m
18	7	3.8	149	2 Q9ZCK5	Q9ZCK5 rickettsia
19	7	3.8	149	2 Q05444	Q05444 mycobacteri

20	7	3.8	152	10 P93801	P93801 zea mays (m
21	7	3.8	164	2 Q9Z582	Q9Z582 streptomyc
22	7	3.8	166	1 Q9YCD6	Q9YCD6 aeropyrum p
23	7	3.8	182	2 Q9KSV2	Q9KSV2 vibrio chol
24	7	3.8	191	1 Q9DMQ3	Q9DMQ3 pyrococcus
25	7	3.8	204	10 Q9LVA6	Q9LVA6 arabidopsis
26	7	3.8	218	2 Q51920	Q51920 prochloroco
27	7	3.8	225	5 Q61752	Q61752 caenorhabd
28	7	3.8	251	2 Q9K3Y4	Q9K3Y4 streptomyc
29	7	3.8	255	10 Q9W9D1	Q9W9D1 arabidopsis
30	7	3.8	260	11 Q9JYJ0	Q9JYJ0 mus musculu
31	7	3.8	268	4 Q9NS92	Q9NS92 homo sapien
32	7	3.8	275	1 Q54485	Q54485 staphylothe
33	7	3.8	295	10 Q22158	Q22158 arabidopsi
34	7	3.8	298	2 Q9RM95	Q9RM95 bradyrhizob
35	7	3.8	300	2 Q52615	Q52615 proteus vul
36	7	3.8	319	2 P96845	P96845 mycobacteri
37	7	3.8	324	2 Q9Z178	Q9Z178 lactococcus
38	7	3.8	325	10 Q9ZRC5	Q9ZRC5 glycine max
39	7	3.8	332	2 Q87803	Q87803 pseudomonas
40	7	3.8	345	10 Q9T024	Q9T024 arabidopsis
41	7	3.8	346	10 Q9SQH3	Q9SQH3 brassica na
42	7	3.8	349	11 Q9JYJ1	Q9JYJ1 mus musculu
43	7	3.8	354	2 Q9KW67	Q9KW67 staphylococ
44	7	3.8	368	11 Q9JYJ2	Q9JYJ2 mus musculu
45	7	3.8	379	11 Q9JYJ3	Q9JYJ3 mus musculu

ALIGNMENTS

RESULT 1
ID Q9M8S8 PRELIMINARY: PRT: 271 AA.
AC Q9M8S8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE MTO-INOSITOL MONOPHOSPHATASE.
GN F13E7.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Bernstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F13E7 genomic sequence.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC018363; AAF26973.1; -
SQ SEQUENCE 271 AA: 29121 MW: 2785848F541B8F5F CRC64;

Query Match 4.4%; Score 8; DB 10; Length 271;
Best local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 IDAKKAG 171
Db 15 IDAKKAG 22

RESULT 2
ID Q9Z8K0 PRELIMINARY: PRT: 366 AA.
AC Q9Z8K0;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE DNA POLYMERASE III (BETA CHAIN).
 GN DNAN.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CWL029;
 RX MEDLINE-99206606; PubMed-10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 RL Nat. Genet. 21:385-389(1999).
 DR EMBL: AE001618; AAD18487.1; -.
 DR INTERPRO: IPR001001; -.
 DR PRAM: PR00712; DNA_pol3_beta: 1.
 SQ SEQUENCE 366 AA; 40334 MW; 8A6D188DB26BC5E1 CRC64;

Query Match 4.4%; Score 8; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82
 |||||
 Db 149 ESRVLTG 156

RESULT 3
 O9PKW4 PRELIMINARY; PRT; 366 AA.
 AC O9PKW4;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE DNA POLYMERASE III, BETA SUBUNIT.
 GN TC0347.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGG;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39";
 RN Nucleic Acids Res. 28:1397-1406(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
 RA Hickey E.K., Peterson J., Umayam L.A., Utterback T., Berry K.,
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
 RA Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
 RA Salzberg S.L., Eisen J., Fraser C.M.;
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003301; AAF39208.1; -.
 DR TIGR: TC0347; -.
 DR INTERPRO: IPR001001; -.
 DR PRAM: PR00712; DNA_pol3_beta: 1.
 SQ SEQUENCE 366 AA; 40493 MW; E3887C463C442D72 CRC64;

Db 149 ESRVLTG 156

RESULT 4
 O9K274 PRELIMINARY; PRT; 366 AA.
 AC O9K274;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE DNA POLYMERASE III, BETA SUBUNIT.
 GN CP0419.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39";
 RN Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE002203; AAF38262.1; -.
 DR TIGR: CP0419; -.
 SQ SEQUENCE 366 AA; 40440 MW; EC54D84CF4085150 CRC64;

Query Match 4.4%; Score 8; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82
 |||||
 Db 149 ESRVLTG 156

RESULT 5
 O9JSG3 PRELIMINARY; PRT; 366 AA.
 AC O9JSG3;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE DNA POLYMERASE III (BETA CHAIN).
 GN DNAN.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J138;
 RX MEDLINE-20330349; PubMed-10871362;
 RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AP002546; BAA96548.1; -.
 DR TIGR: CP0419; -.
 SQ SEQUENCE 366 AA; 40364 MW; E8FA92ACF76BD1B0 CRC64;

Query Match 4.4%; Score 8; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82
 |||||
 Db 149 ESRVLTG 156

Query Match 4.4%; Score 8; DB 2; Length 366;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

75 ESRVLTG 82
 |||||


```

RESULT 6
ID 084078 PRELIMINARY; PRT; 416 AA.
AC 084078;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE DNA POL III (BETA CHAIN).
GN DNAN.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL; AE001282; AAC67666.1; -.
DR HSSP; P00583; 2POL.
DR INTERPRO; IPR001001; -.
DR PRAM; PF00712; DNA_pol3_beta; 1.
SQ SEQUENCE 416 AA; 46529 MW; F32BD087A8930683 CRC64;

```

```

Query Match 4.4%; Score 8; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 75 ESRVLTG 82
Db 199 ESRVLTG 206

```

```

RESULT 7
ID 099006 PRELIMINARY; PRT; 428 AA.
AC 099006;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ENDOCHITINASE.
OS Trichoderma hamatum.
OC Eukaryota; Fungi; Ascomycota; anamorphic Ascomycota; Trichoderma.
OX NCBI_TaxId=49224;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAM-61;
RX MEDLINE=97132641; PubMed=8978093;
RA Pekete C., Weszely T., Hornok L.;
RT "Assignment of a PCR-amplified chitinase sequence cloned from
RT Trichoderma hamatum to resolved chromosomes of potential biocontrol
RT species of Trichoderma."
RL FEMS Microbiol. Lett. 145:385-391(1996).
DR EMBL; Z71415; CAA96021.1; -.
DR HSSP; P07254; 1CTN.
DR INTERPRO; IPR001223; -.
DR INTERPRO; IPR001579; -.
DR PRAM; PF00704; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KM Hydrolase; Glycosidase.
SQ SEQUENCE 428 AA; 46881 MW; 0901641A987D8A59 CRC64;

```

```

Query Match 4.4%; Score 8; DB 3; Length 428;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 159 PSASIDA 166
Db 139 PSASIDA 146

```

```

RESULT 8
ID 09RT28 PRELIMINARY; PRT; 470 AA.
AC 09RT28;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN DRI599.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxId=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the Radioresistant Bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
DR EMBL; AE002004; AAF11162.1; -.
DR TIGR; DRI599; -.
DR INTERPRO; IPR000126; -.
DR PRINTS; PR00839; V8PROTEASE.
SQ SEQUENCE 470 AA; 50683 MW; 60D8581386287562 CRC64;

```

```

Query Match 4.4%; Score 8; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 57 AGADGALT 64
Db 392 AGADGALT 399

```

```

RESULT 9
ID 09JY67 PRELIMINARY; PRT; 1067 AA.
AC 09JY67;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE MULTIPLE TRANSFERABLE RESISTANCE SYSTEM PROTEIN MTRD.
GN NMB1715.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2015755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Debey R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain

```

RT MC58.";
 RL Science 287:1809-1815(2000).
 DR EMBL: AF002521; AAF42062.1; -.
 DR TIGR: NMB1715; -.
 SQ SEQUENCE 1067 AA; 113973 MW; A05545371B7943C2 CRC64;

Query Match
 Best Local Similarity 4.4%; Score 8; DB 2; Length 1067;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAEAGITG 43
 |||||
 Db 1054 AAEAGITG 1061

RESULT 10
 O9JUT51 PRELIMINARY; PRT; 1067 AA.

ID O9JUT51
 AC O9JUT51;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE PROBABLE DRUG EFFLUX PROTEIN.
 GN MTRD OR NMA1969.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=65699;

RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
 Rajagaram M.A., Rutherford K.M., Mungall K., Quail M.A.,
 Whitehead S., Spratt B.G., Barrall B.G.;
 RT *Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491.";
 RL Nature 404:502-506(2000).
 DR EMBL: AL162757; CAB85189.1; -;
 SQ SEQUENCE 1067 AA; 114012 MW; D3216DD7F4783B41 CRC64;

Query Match
 Best Local Similarity 4.4%; Score 8; DB 2; Length 1067;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAEAGITG 43
 |||||
 Db 1054 AAEAGITG 1061

RESULT 11

O9WMP8 PRELIMINARY; PRT; 86 AA.

ID O9WMP8
 AC O9WMP8;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE HRCS.
 GN HRCS.

OS Xanthomonas campestris (pv. vesicatoria).
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 NCBI_TaxID=341;

OX NCB1_TaxID=341;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=75-3;
 RX MEDLINE=98453137; PubMed=9781876;
 RA Huguet E., Hahn K., Wengelnik K., Bonas U.;
 RT *DppA mutants of Xanthomonas campestris pv. vesicatoria are affected

RT in pathogenicity but retain the ability to induce host-specific
 RT hypersensitive reaction.";
 RL Mol. Microbiol. 29:1379-1390(1998).
 DR EMBL: AF056246; AAD21322.1; -.
 DR INTERPRO: IPR002191; -.
 DR PFAM: PF01313; Bac_export_3; 1.
 DR PRINTS: PR00952; TYPE3IMOPROT.
 SQ SEQUENCE 86 AA; 8914 MW; CD44BF70739BAD82 CRC64;

Query Match
 Best Local Similarity 3.8%; Score 7; DB 2; Length 86;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 VVAIAIV 11
 |||||
 Db 58 VVAIAIV 64

RESULT 12
 O9RBR3 PRELIMINARY; PRT; 86 AA.

ID O9RBR3
 AC O9RBR3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE HRCS HOMOLOG.
 OS Xanthomonas campestris (pv. glycines).
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 NCBI_TaxID=36865;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8RA;
 RA Oh C., Heu S., Yoo J.Y., Cho Y.S.;
 RT "A hrcV-homologous gene mutant of Xanthomonas campestris pv. glycines
 RT 8ra that lost pathogenicity on the host plant but was able to elicit
 RT the hypersensitive response on nonhosts.";
 RL Mol. Plant Microbe Interact. 12:633-639(1999).
 DR EMBL: AF160974; AAD46903.1; -.
 DR INTERPRO: IPR002191; -.
 DR PFAM: PF01313; Bac_export_3; 1.
 DR PRINTS: PR00952; TYPE3IMOPROT.
 SQ SEQUENCE 86 AA; 8928 MW; DA4FD2D1A39BAD82 CRC64;

Query Match
 Best Local Similarity 3.8%; Score 7; DB 2; Length 86;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 VVAIAIV 11
 |||||
 Db 58 VVAIAIV 64

RESULT 13

O9K363 PRELIMINARY; PRT; 86 AA.

ID O9K363
 AC O9K363;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HRCS (HRP3).
 GN HRCS.

OS Xanthomonas oryzae pv. oryzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 NCBI_TaxID=64187;

OX NCB1_TaxID=64187;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFP 311018;
 RA Ochiai H.;

RT *Construction and characterization of a Xanthomonas oryzae pv. oryzae

RT bacterial artificial chromosome library."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF301237;
 RA Oku T.;
 RT "Clustered hrp genes in Xanthomonas oryzae pv. oryzae."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB045311; BAB07863.1; -;
 DE EMBL: AB040134; BAA92827.1; -;
 SQ SEQUENCE 86 AA; 8944 MW; 3844AA65739BAD98 CRC64;

Query Match 3.8%; Score 7; DB 2; Length 86;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VVAIAV 11
 DB 58 VVAIAV 64

RESULT 14
 ID Q9MCN5 PRELIMINARY; PRT; 106 AA.
 AC Q9MCN5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HOLIN.
 GN 70.
 OS Bacteriophage HK97.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=37554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Juhala R., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,
 RT Hendrix R.W.;
 RT "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic
 RT mosaicism in the lambdaoid bacteriophages."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF069529; AAF31096.1; -;
 SQ SEQUENCE 106 AA; 11274 MW; CA91992E1C30769E CRC64;

Query Match 3.8%; Score 7; DB 9; Length 106;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAKKAGV 172
 DB 94 AAKKAGV 100

RESULT 15
 ID Q9MCT3 PRELIMINARY; PRT; 107 AA.
 AC Q9MCT3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HOLIN.
 OS Bacteriophage HK022.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10742;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Juhala R., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,
 RA Hendrix R.W.;
 RT "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic
 RT mosaicism in the lambdaoid bacteriophages.";

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF069308; AAF30386.1; -;
 SQ SEQUENCE 107 AA; 11340 MW; 18A92A2A20FDICAB CRC64;

Query Match 3.8%; Score 7; DB 9; Length 107;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAKKAGV 172
 DB 95 AAKKAGV 101

Search completed: April 6, 2001, 01:28:06
 Job time: 920 sec

-

vector (pBluescript II SK+) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=56) and sequence (n=16). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 65 a 78 c 83 g 38 t

alignment_scores: Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x A1973896 ..

Align seg 1/1 to: A1973896 from: 1 to: 264

166 AAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
167 GCGGCCAGAGAGCGCGGTGACACAC 193

seq_name: gb_gss14:AQ998089

seq_documentation_block:

LOCUS AQ998089 439 bp DNA GSS 24-FEB-2000
DEFINITION RPCI-23-279B21.TV RPCI-23 Mus musculus genomic clone RPCI-23-279B21
, DNA sequence.

ACCESSION AQ998089
VERSION AQ998089.1 GI:7073186

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 439)

AUTHORS Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akınret, B., Levin, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C. M.

JOURNAL Mouse BAC End Sequences from Library RPCI-23
MEDLINE Unpublished (1999)

COMMENT Other GSSs: RPCI-23-279B21.TV

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resear h Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac.ends/mouse/bac_end_intro.html

Plate: 279 row: B column: 21
Seq primer: SP6

Class: BAC ends.
Location/Qualifiers
1. 439

FEATURES
source

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-279B21"
/clone_id="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 152 a 84 c 104 g 99 t

alignment_scores: Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x AQ998089 ..

Align seg 1/1 to: AQ998089 from: 1 to: 439

75 GUserArgTyrrValIleuThrGlyArg 83
|||||
158 GAAGACAGGTATGCTCTCAGTGGAGGA 184

seq_name: gb_gss12:AQ836896

seq_documentation_block:

LOCUS AQ836896 574 bp DNA GSS 30-AUG-1999
DEFINITION HS_5463_A1_H09.T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-1039 Col=17 Row=O, DNA sequence.

ACCESSION AQ836896
VERSION AQ836896.1 GI:5806770

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 574)

AUTHORS Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D., and Hood, L.

JOURNAL Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
MEDLINE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

COMMENT 99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 1039 row: O column: 17
Seq primer: T7
High quality sequence stop: 574.

Class: BAC ends
Location/Qualifiers
1. 574

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1039 Col=17 Row=O"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="Male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 220 a 87 c 109 g 149 t 9 others
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x AQ836896/rev ..

Align seg 1/1 to reverse of: AQ836896 from: 1 to: 574

10 AlavalserrathrthvalSerile 18
|||||
185 GCTGTCACTGCTACTGCTCTAT 159

seq_name: gb_est5:BE726886

seq_documentation_block:

LOCUS BE726886 606 bp mRNA EST 14-SEP-2000
DEFINITION 894096810.y2 C. reinhardtii CC-1690, normalized, lambda zap II
ACCESSION BE726886
VERSION BE726886.1 GI:10128182
KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 606)

REFERENCE Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; Project phase 2
Unpublished (2000)

JOURNAL CONTACT: Charles Hauser
COMMENT DCMB Box 91000

Duke University
Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

FEATURES Location/Qualifiers

1..606 /organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, normalized, lambda zap II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in 7AP (acetate-containing) medium in the light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

light, 7AP medium in the dark, HS (minimal) medium in

US-09-589-870-2 x BE726886 ..

Align seg 1/1 to: BE726886 from: 1 to: 606

164 lIleaspAlaAlaLysLysAlaGlyVal 172
|||||
468 ATCGATGCTGCCAGAGCGCGCGTC 494

seq_name: gb_est11:AA749932

seq_documentation_block:

LOCUS AA749932 184 bp mRNA EST 20-JAN-1998
DEFINITION ISAS0492 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
CDNA clone ISAS0492, mRNA sequence.

ACCESSION AA749932
VERSION AA749932.1 GI:2796638
KEYWORDS EST.

SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

1 (bases 1 to 184)

REFERENCE Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee

,M.C. and Eun,M.Y.

Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

Unpublished (1998)

JOURNAL CONTACT: Eun M.Y.

Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA

Suwon, Kyungdo, Korea

Tel: 82 331 290 0301

Fax: 82 331 290 0307

Email: myeunesun20@sl.re.kr

Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji

University, Yongin, Korea. 449-728 bhna@bioserver.myongji.ac.kr

Seq primer: M3 Reverse Primer.

FEATURES Location/Qualifiers

1..184

/organism="Oryza sativa"

/cultivar="Milyang23"

/db_xref="taxon:4530"

/clone="ISAS0492"

/clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"

/tissue_type="Immature Seed"

/dev_stage="5 days after pollination"

/lab_host="E. coli SOLR"

/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional CDNA library inserted into lambda ZAPII

vector at 5' end with EcoRI and 3' end with Xho I site."

BASE COUNT 31 a 79 c 29 g 45 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x AA749932 ..

Align seg 1/1 to: AA749932 from: 1 to: 184

20 AlaserrAlaAspProSer 27
|||||

69 GCTCGCGCTCCGCCGATCATCC 92

seq_name: gb_est50:AW894265

seq_documentation_block:

LOCUS AW894265 223 bp mRNA EST 24-MAY-2000
DEFINITION CM3-NN0030-110500-178-fil NN0030 Homo sapiens CDNA, mRNA sequence.

ACCESSION AM894265
 VERSION AM894265.1 GI:8058470
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 223)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ft2-CM3-NN0030-110500-178-11&f3=2000-05-11&f4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 22
 High quality sequence stop: 77.
 Location/Qualifiers
 1..223
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0030"
 /dev_stage="Adult"
 /note="Organ: nervous.normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 62 a 53 c 39 g 69 t
 ORIGIN
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-589-870-2 x AM894265 ..
 Align seg 1/1 to: AM894265 from: 1 to: 223
 seq_name: gb_est16:AI084542
 seq_documentation_block: 258 bp mRNA EST 27-AUG-1998
 LOCUS AI084542
 DEFINITION OV43004.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1640071
 3', mRNA sequence.
 ACCESSION AI084542
 VERSION AI084542.1 GI:3422965
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 258)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo Ph.D.
 ' cDNA library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 900 Std Error: 0.00
 Seq primer: -40m3 fwd. ET from Amersham
 High quality sequence stop: 221.
 Location/Qualifiers
 1..258
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:1640071"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories Inc. and primed with a Not I - oligo(dT) primer [5' TGTTCACATCTGAAGTGGAGCGGCCGCCCAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 62 a 73 c 77 g 46 t
 ORIGIN
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-589-870-2 x AI084542/rev ..
 Align seg 1/1 to reverse of: AI084542 from: 1 to: 258
 seq_name: gb_est24:AI756574
 seq_documentation_block: 299 bp mRNA EST 18-JAN-2000
 LOCUS AI756574
 DEFINITION E5E5Ea02f10.y1 Elmeria M5-6 Merozoite stage Elmeria tenella cDNA
 5', mRNA sequence.
 ACCESSION AI756574
 VERSION AI756574.1 GI:5150297
 KEYWORDS EST.
 SOURCE Elmeria tenella.
 ORGANISM Elmeria tenella
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriidae; Elmeriidae;
 REFERENCE 1 (bases 1 to 299)
 Liberators, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,

TITLE
JOURNAL
COMMENT
Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Glibbons, M., Pape, D., Harvey, N., Schurt, R., Riltter, E., Koh, S., Florence, N., Shin, T., Jackson, Y., Cardenas, M., Mcann, R., Waterston, R., Wilson, R. and Sibley, D. WashU-Merck Elmeria tenella project
Unpublished (1999)
Contact: David Sibley, Ph.D.
WashU-Merck Elmeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevatson.wustl.edu
Contact David Sibley (toxeast@orcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Glibco
High quality sequence stop: 298.
Location/Qualifiers

FEATURES

source

1. 299
/organism="Elmeria tenella"
/strain="IS18"
/db_xref="taxon:5802"
/clone_lib="Elmeria M5-6 Merozoite stage"
/dev_stage="Merozoite"
/lab_host="SOLR E. coli"
/note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI
; Merozoites were obtained from ceecal scrapings of chickens infected with E. tenella. The library may contain a small percentage of host or bacterial contaminants. CDNA was synthesized from poly RNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on Sephacryl S500. cDNAs were digested with EcoRI/XhoI and cloned into lambda Zap II (Stratagene). Clones were converted to phagemids by mass excision using Exassist helper phage and SOLR cells (Stratagene). Insert sizes range from 0.7-1.5 kb."
BASE COUNT 38 a 103 c 84 g 73 t 1 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-589-870-2 x A1756574/rev ..

Align seg 1/1 to reverse of: A1756574 from: 1 to: 299

32 A1aG1nValSerAlaAlaG1uAla 39
|||||
77 GCACAGGTTCTGCTGCTGAGGCT 54

seq_name: gb_est17:A1194189

seq_documentation_block:
LOCUS A1194189 311 bp mRNA EST 13-OCT-1998
DEFINITION ue81c04.r1 Soares_NMPu Mus musculus cDNA clone IMAGE:1497510 5',
RNA sequence.
ACCESSION A1194189
VERSION A1194189.1 GI:3745396
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 311)
Marfa, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

TITLE
JOURNAL
COMMENT
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Wmouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:935114
Seq primer: -28mJ rev2 ET from Amersham
High quality sequence stop: 213.
Location/Qualifiers

FEATURES

source

1. 311
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="IMAGE:1497510"
/clone_lib="Soares_NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pregnant mouse uterus, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 86 a 52 c 62 g 111 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-589-870-2 x A1194189 ..

Align seg 1/1 to: A1194189 from: 1 to: 311

12 serLeuThrValSerIleThr 19
|||||
115 TCACCTACACTGTCAGTATTACC 138

seq_name: gb_est37:AV540058

seq_documentation_block:
LOCUS AV540058 326 bp mRNA EST 07-SEP-2000
DEFINITION AV540058 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
CDNA clone R2143h06f 3', mRNA sequence.
ACCESSION AV540058
VERSION AV540058.1 GI:8701816
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS
TITLE
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
JOURNAL
MEDLINE
COMMENT
Contact: Erika Asamizu

REFERENCE 1 (bases 1 to 373)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watsn.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 295.
 Location/Qualifiers
 source 1..373
 /organism="Homo sapiens"
 /db_xref="GDB:3807716"
 /db_xref="taxon:9606"
 /clone="IMAGE:502728"
 /clone_1lb="Soares_pregnant_uterus_NbHPV"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73-Pac; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5', AACGTGAGAGATTCGCGCGCCGCTTTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
 BASE COUNT 87 a 89 c 98 g 94 t 5 others
 ORIGIN
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-589-870-2 x AA126930 ..
 Align seg 1/1 to: AA126930 from: 1 to: 373
 131 GlnTrrpleuthrSerglyThr 138
 ||||||||||||||||||
 320 CAGTGCCTCTGACTTCAGGACN 343
 seq_name: gb_gss23:B40058
 seq_documentation_block:
 LOCUS B40058 374 bp DNA GSS 18-OCT-1997
 DEFINITION HS-1050-B2-F08-WF-ab1 CIT Human Genomic Sperm Library C Homo
 sapiens genomic clone Plate=CT 772 Col=16 Row=L, DNA sequence.
 ACCESSION B40058
 VERSION B40058.1 GI:2544310
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 374)
 AUTHORS Mahairas, G.G., Zackrone, K.D., Smith, T., Tipson, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
 TITLE Construction of a Characterized Clone Resource for Genomic

JOURNAL Unpublished (1997)
 COMMENT Contact: Mahairas GC, Zackrone KD, Hood L
 University of Washington
 Seattle, WA 98195, USA
 Tel: (206) 616-8744
 Fax: (206) 685-7301
 Email: kzackrone@u.washington.edu
 Sequence Tagged Connector
 Plate: CT 772 row: L column: 16
 Class: BAC ends
 High quality sequence stop: 374.
 Location/Qualifiers
 source 1..374
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=CT 772 Col=16 Row=L"
 /clone_1lb="CIT Human Genomic Sperm Library C"
 /sex="M"
 /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 111 a 66 c 78 g 119 t
 ORIGIN
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-589-870-2 x B40058/rev ..
 Align seg 1/1 to reverse of: B40058 from: 1 to: 374
 133 leuthrSerglyThrtrglu 140
 ||||||||||||||||||
 368 CTACTCAGATCAGGACCGGAA 345
 seq_name: gb_gss16:AZ131983
 seq_documentation_block:
 LOCUS AZ131983 377 bp DNA GSS 02-JUN-2000
 DEFINITION OSJNB0111P08r CUGI Rice BAC Library (ECORI) Oryza sativa genomic
 clone OSJNB0111P08r, DNA sequence.
 ACCESSION AZ131983
 VERSION AZ131983.1 GI:8210478
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
 REFERENCE 1 (bases 1 to 377)
 AUTHORS Wing, R.A. and Dean, R.A.
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
 JOURNAL Unpublished (1998)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: CGAAGACGCTATGACCATG
 Class: BAC ends
 High quality sequence start: 14
 High quality sequence stop: 363.
 Location/Qualifiers
 source 1..377
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"

```

/db_xref="taxon:4530"
/clone_lib="OSJNB011P08-"
/clone_lib="CUCI Rice BAC Library (ECORI)"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

BASE COUNT 87 a 71 c 82 g 136 t 1 others

ORIGIN

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:
US-09-589-870-2 x AZ131983/rev ..

Align seg 1/1 to reverse of: AZ131983 from: 1 to: 377

```

14 ThrThrValSerIleThrAlaSer 21
|||||
360 ACCACCGTGTCTATTACGCGCAGC 337
```

seq_name: gb_estf2:BE553360

seq_documentation_block:

```

LOCUS BE553360 390 bp mRNA 15-AUG-2000
DEFINITION ur45c10.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153234 5',
mRNA sequence.
```

```

ACCESSION BE553360
VERSION BE553360.1 GI:9817847
```

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 390)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:1055990

Seq primer: -40RP from Gibco

High quality sequence stop: 310.

Location/Qualifiers

FEATURES

1..390
/organism="Mus musculus"

```

/strain="FVB-3"
/db_xref="taxon:10090"
/clone_lib="IMAGE:3153234"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
```

BASE COUNT 110 a 68 c 87 g 125 t

ORIGIN

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:
US-09-589-870-2 x BE553360 ..

Align seg 1/1 to: BE553360 from: 1 to: 390

```

12 SerLeuThrThrValSerIleThr 19
|||||
```

```

356 TCACTTACAACTGTGATATTACG 379
```

THIS PAGE BLANK (USPTO)

PT positive bacterial protein.
XX
XX Disclosure; Fig 1b; 54pp; English.
XX
XX Tetrameric biologically active streptavidin is produced by secretion
CC from *Bacillus subtilis* transformed with a plasmid encoding the
CC sequence.
XX
SQ Sequence 183 AA;

Query Match 100.0%; Score 936; DB 14; Length 183;
Best Local Similarity 100.0%; Pred. No. 2.5e-76;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIYVAIAIVSLITVYITASASADPSKDSKAQVSAEAGITGTWYQLGSTFIYTAGAD 60
DB 1 mrkivvaiaivslitvysltasasadpskdskaqvsaeeagltgwyngqlgstflvtagad 60
QY 61 GALTGYESAIVGNAESRYVLTGRYDSAPATDGSCTALGWTVAKNKNNYRNHSAATTWNGOY 120
DB 61 galtgtyesavgnaesryvltgrydsapacdgsgtalgwvawknnyrnhsatcwsqy 120
QY 121 VGGAEARINTQWLLTSGTTEANAMKSTLVGHDPFTTKVYKPSAASIDAAKKAGVNNGNPLDA 180
DB 121 vggaearinrcqwltsgtteanawkstlvghdfttkvypsaasidaakkagvnngnplda 180
QY 181 VQO 183
DB 181 vqg 183

RESULT 4
W29306 ID W29306 standard; Protein; 183 AA.
XX
AC W29306;
XX
DT 27-APR-1998 (first entry)
XX
DE Wild-type streptavidin protein.
XX
KM Streptavidin; biotin; anti-interference reagent; detection; mutain;
KW avidin; non-specific binding.
XX
OS Unidentified.
XX
XX
XX Key Location/Qualifiers
FH Peptide 1..24
FT /label= signal
FT Protein 25..183
XX
XX DE19637718-A1.
XX
XX
XX 02-OCT-1997.
XX
XX 16-SEP-1996; 96DE-1037718.
XX
XX 01-APR-1996; 96DE-1013053.
XX
XX (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
XX Brandstetter H, Deger A, Engh R, Kopetzki E, Mueller R;
PI Schmitt U;
XX
XX WPI; 1997-482043/45.
XX
XX N-PSDB; T73193.
XX
XX Streptavidin and avidin mutants with reduced binding affinity for
PT biotin - useful for reducing interference from nonspecific binding
PT in assays
XX
XX Disclosure; Page 17-18; 26pp; German.

XX
XX This sequence represents a streptavidin which is used in a novel method
CC of reducing interference from non-specific binding in assays. Mutains
CC constructed from a core streptavidin or avidin sequence are selected that
CC differ from the native polypeptide by at least one amino acid and have a
CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable
CC polypeptide may be present as a polymeric conjugate, e.g. with another
CC polypeptide or protein, especially bovine serum albumin. These mutants
CC are used as anti-interference reagents for reducing and/or avoiding
CC nonspecific interactions in a process for detecting an analyte. In
CC particular, they are used in assays where the streptavidin/avidin-biotin
CC specific binding pair is involved for qualitative and/or quantitative
CC determination of an analyte in a test sample, e.g. a heterogeneous
CC immunoassay or a hybridisation assay. Despite having a lower binding
CC affinity for biotin, the mutants have high immunological cross-reactivity
CC with native streptavidin and avidin.
XX
SQ Sequence 183 AA;

Query Match 100.0%; Score 936; DB 18; Length 183;
Best Local Similarity 100.0%; Pred. No. 2.5e-76;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIYVAIAIVSLITVYITASASADPSKDSKAQVSAEAGITGTWYQLGSTFIYTAGAD 60
DB 1 mrkivvaiaivslitvysltasasadpskdskaqvsaeeagltgwyngqlgstflvtagad 60
QY 61 GALTGYESAIVGNAESRYVLTGRYDSAPATDGSCTALGWTVAKNKNNYRNHSAATTWNGOY 120
DB 61 galtgtyesavgnaesryvltgrydsapacdgsgtalgwvawknnyrnhsatcwsqy 120
QY 121 VGGAEARINTQWLLTSGTTEANAMKSTLVGHDPFTTKVYKPSAASIDAAKKAGVNNGNPLDA 180
DB 121 vggaearinrcqwltsgtteanawkstlvghdfttkvypsaasidaakkagvnngnplda 180
QY 181 VQO 183
DB 181 vqg 183

RESULT 5
W59216 ID W59216 standard; Protein; 183 AA.
XX
AC W59216;
XX
XX 27-AUG-1998 (first entry)
XX
XX S. avidinli streptavidin protein.
XX
DE Streptavidin; ligand; binding affinity; mutant; isolation;
KW purification; recover; immobilise.
XX
XX Streptomycetes avidinli.
XX
XX OS
XX PN EP835934-A2.
XX
XX 15-APR-1998.
XX
XX 09-OCT-1997; 97EP-0117504.
XX
XX 10-OCT-1996; 96DE-1041876.
XX
XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX Skerra A, Voss S;
XX
XX WPI; 1998-218868/20.
XX
XX N-PSDB; V34714.
XX
XX Streptavidin mutants with higher binding affinity for peptide
PT ligands - have mutation in amino acid region 44-53, used to isolate,
PT

PT purify or determine fusion proteins including these ligands
XX
PS Disclosure; Page -; 21pp; German.
XS
CC This sequence encodes a wild-type streptavidin protein isolated from
CC Streptomyces avidinii. This sequence is used to produce mutants which
CC are used in a method to assay the binding affinity of streptavidin
CC mutants. These mutants have a mutation within the amino acid (aa) region
CC 44-53 of the wild-type protein show a higher binding affinity than the
CC wild-type for peptide ligands that include the sequence of formula
CC Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly,
CC or Y = Gln and Z = Arg or Lys. Recombinant streptavidin mutants can
CC be used to isolate, purify and determine proteins or to determine/recover
CC substances that contain streptavidin-binding groups. Such compounds may
CC also be used to immobilise fusions on microtitre plates, microbeads or
CC sensor chips.
CC NOTE: This sequence does not appear in the specification but is used to
CC make the mutant streptavidin proteins represented in W59Z17 and W59Z18.
XX
XQ Sequence 183 AA;

Query Match	100.0%	Score 936	DB 19	Length 183
Best Local Similarity	100.0%	Pred. No. 2,5e-76		
Matches 183	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MRIIVAAIAVSLTIVTSITASASADPSKDSKNOVSAAEAGITGTYWYNOLGSPFTITAGAD	60	
Db	1	mrkIIVaaIAVslTlvtSIItasasadpSkdsKqVsaAaeagItgTywnqLgSfItItagad	60	
QY	61	GALGTYSAAVGNMESRVLVLTGRDYSAAITDSSGALGTYVAAKKNRYRNAHSATTWSGCY	120	
Db	61	galgtYsaAvgnaesrVlvtlgYysadpTgsgtAlgtVaaKknYrnahsatttwsgcy	120	
QY	121	VGGAEARINTOMLTSTGTTEANAMKSTIVGHDFTFKVSPASIDAAKKAGVNNCPIDA	180	
Db	121	vggEarIntgWlTstgtTeaNaMkStIvghDfTfKvSpaSiDaakKagvnnCpIda	180	
QY	181	VQQ 183		
Db	181	vqQ 183		

RESULT	6
ID	Y17868
XX	
XX	Y17868 standard; Protein; 183 AA.
AC	Y17868;
XX	
DT	20-AUG-1999 (first entry)
XX	
DE	Streptococcus streptavidin.
XX	
KM	Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;
XX	hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin.
OS	Streptococcus sp.
XX	
PN	MO9929838-A1.
PD	17-JUN-1999.
XX	
PF	09-DEC-1998; 98WO-US26086.
XX	
PR	09-DEC-1997; 97US-0067978.
XX	
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.
PI	Cederholm-Williams SA;
DR	WPI; 1999-385599/32.
XX	
DR	N-PSDB; X80198.
XX	

PT A fibrinogen-converting enzyme fusion protein
XX
PS Disclosure: Page 28; 35pp; English.
XX
CC The present invention describes a fibrinogen-converting enzyme fusion
CC protein (FCE). The fusion protein is a multidomain protein comprising:
CC (a) a FCE; and (b) a first member of a binding pair, that is linked to
CC the RCE chain; (i) directly by bonds utilizing the N-terminal amino
CC groups, the C-terminal carboxy groups or side-chain functionalities;
CC (ii) via a bifunctional linkage moiety linking the groups or
CC functionalities; or (iii) by the first member binding to the second
CC member of the binding pair, where the second member of the binding pair
CC is covalently attached to the first polypeptide chain. The FCE can be
CC used in a method for producing fibrin. Fibrin is useful as a sealant in
CC surgery to, e.g., reduce bleeding by sealing blood vessels, and tissues
CC that have been dissected either in surgery or through wounding. The
CC fusion protein allows for the removal of the fibrinogen converting
CC enzyme from the fibrin sealant preparation via the binding of
CC streptavidin to a biotin solid support. The present sequence represents
CC Streptococcus streptavidin as given in the present invention.

Query Match	100.0%	Score 936:	DB 20:	Length 183:
Best Local Similarity	100.0%	Pred. NO. 2.5e-76:		
Matches 183:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	1	MKKIIVAAIAVSLTVSTITASADPDSKSKQVAAEAGITGTVYNOLSGSTFIYTAGAD	60	
Db	1	mrkiivaaiaavsltlvtstltasasadpskskqvaasaaegitgtvynqigstfitytagad	60	
QY	61	GALTITYESAVGNAASRYLTGRTDSPAFTDSSGTLALGTVAMKKNRYRNAHSATTWSGOY	120	
Db	61	galitltesaavgaesryvltgrytsapatdsgtcalgtvawkmnyrnabattwsqgy	120	
QY	121	VGGARINTQWLTFSGTTEANAMKSTIVGHPTFTKVPSPASIDAAKKAGVNCNPDLDA	180	
Db	121	vggaearintqwltsqgteanawkstlvgnhtftkvpspaasidaakkagvnmnplda	180	
QY	181	VQO 183		
Db	181	vqg 183		

RESULT	7	
Y44701		
ID	Y44701	standard; Protein; 183 AA.
XX		
AC	Y44701;	
XX		
DT	25-APR-2000	(first entry)
XX		
DE	Streptavidin protein for recombinant PART27 vector.	
XX		
KW	Potato protease inhibitor-II; PPI-II; streptavidin; worm;	
KW	insect; plant-toxicous protein; pest resistance; moth; insect; weevil;	
KW	grub; beetle; fly; thrip; locust; cricket; borer; mte; looper;	
KW	insecticidal.	
XX		
OS	Unidentified	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..24
FT		/note= "Signal peptide"
XX		
PN	W0200004049-A1.	
XX		
PD	27-JAN-2000.	
XX		
PF	15-JUL-1999;	99WO-NZ00110.
XX		
PR	15-JUL-1998;	98NZ-0331002.

XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
PA Christeller JF, Sutherland PM, Murray C, Markwick NP, Phillip BA;
XX Malone LA, Burgess EPU;
PI WPI: 2000-171244/15.
XX N-PSDB: 249867.
DR
XX New chimeric polypeptide and composition comprising the polypeptide
PT useful for conferring pest resistance on plants -
XX
XX Disclosure; Fig 12; 11pp; English.
XX
XX The present sequence is streptavidin, a plant-toxicous protein.
CC Recombinant vector, PART7 expressing a chimeric polypeptide comprising
CC streptavidin mature peptide fused to the potato proteinase inhibitor-II
CC (PPI-II) signal peptide is targeted to the vacuole.
CC Transformation of plant genome with the vector can produce pest
CC resistance in plants, plant derived products and stored harvest
CC material. Pests that can be controlled include; cotton bollworm,
CC tropical army-worm, European corn-borer or red mite, tobacco horn worm,
CC loopers, rice stem borer, portia, cutworms, diamondback moth, potato
CC tuber moth, codling moth, indian meal moth, gypsy moth, argentine stem
CC weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat
CC weevils, mealworms, flour beetles, black field cricket, locusts,
CC sawflies, Western flower thrips, Hessian flies or two-spotted mite.
XX
SQ Sequence 183 AA:

Query Match 100.0%; Score 936; DB 21; Length 183;

Best Local Similarity 100.0%; Pred. No. 2.5e-76;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRIIVAAIAVSLTTVSTITASASADPSKSKAQAASAEGITGTWYNOLGSTRFTVYAGAD 60
DB 1 mklivvaalavslitvtasasaadpskksaqvsaagltgtwnglqgstltvtagad 60
OY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNVRNAHSATTWSGCY 120
DB 61 galgtgyesaavgnaesryvltgrysdapatdgsctalgtwvkaaknnyrnhsattwsgcy 120
OY 121 VGGAERINTOWLLTSGTTEANAMKSTLVGHDFTFTVKPSASIDAAKKGAVNNGNPIDA 180
DB 121 vggaearrintowlltsgtteanawkstlvghdftfkvpksaasidaakkagvnnnpida 180
OY 181 VQG 183
DB 181 vqg 183

RESULT 8

Y80512 Y80512 standard; Protein: 186 AA.

AC Y80512;

DT 06-JUN-2000 (first entry)

DE Streptomyces avidinl1 sps protein.

XX Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; sps gene; plant development; plant morphology; flower;
KM fruit plant.

XX Streptomyces avidinl1.

OS WO200007427-A2.

PN 17-FEB-2000.

XX 30-JUL-1999; 99WO-IL00420.

XX 03-AUG-1998; 98IL-0125632.
XX
XX (AGRI-) AGRIC RES ORG.
XX
XX Kapulnik Y, Ginzberg I;
XX
XX WPI: 2000-195402/17.
XX N-PSDB: 291073.
DR
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
XX Examples; Page 84; 91pp; English.
PS
XX
XX

CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinl1 sps protein as an example of a heterologous protein introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 186 AA:

Query Match 100.0%; Score 936; DB 21; Length 186;

Best Local Similarity 100.0%; Pred. No. 2.5e-76;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRIIVAAIAVSLTTVSTITASASADPSKSKAQAASAEGITGTWYNOLGSTRFTVYAGAD 60
DB 4 mklivvaalavslitvtasasaadpskksaqvsaagltgtwnglqgstltvtagad 63
OY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNVRNAHSATTWSGCY 120
DB 64 galgtgyesaavgnaesryvltgrysdapatdgsctalgtwvkaaknnyrnhsattwsgcy 123
OY 121 VGGAERINTOWLLTSGTTEANAMKSTLVGHDFTFTVKPSASIDAAKKGAVNNGNPIDA 180
DB 124 vggaearrintowlltsgtteanawkstlvghdftfkvpksaasidaakkagvnnnpida 183
OY 181 VQG 183
DB 181 vqg 183

RESULT 9

Y80513 Y80513 standard; Protein: 184 AA.

AC Y80513;

DT 06-JUN-2000 (first entry)

DE Streptomyces avidinl1 mst protein.

XX Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; mst gene; plant development; plant morphology; flower;
KM fruit plant.

XX Streptomyces avidinl1.

OS WO200007427-A2.

PN 17-FEB-2000.

```

XX 30-JUL-1999; 99WO-IL00420.
PF
XX 03-AUG-1998; 98IL-0125632.
PR
XX (AGRI-) AGRIC RES ORG.
PA
XX Kapulnik Y, Ginzberg I;
PI
XX WPI: 2000-195402/17.
DR N-PSDB; Z91074.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
XX Examples; Page 85; 91pp; English.
PS
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidin1 mat protein as an example of a heterologous protein introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
XX Sequence 184 AA:
SQ

```

```

Query Match          99.5%; Score 931; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 7e-76;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 RKIVAAIAVSLTTVSTITSASADPSKSKAQYSAEAGITGTWYNQLGSTFIYTAGADG 61
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3 rkivaaiaivslttvstitsasadpskkskaqysaaegltgwyngqlgstfiytagadg 62

QY 62 ALTGYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVAAMKNRYRNAHSATTMSGQYV 121
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 63 altgyesavгнаesryvltgrydsapadgsctalgwtvawkmnyrnahsattmsgqyv 122

QY 122 GGAEARINTQWILTSGTTEANMWSKSLVGHDPFTTKVPSAASIDAAKKAGVNNGNPLDAY 181
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 123 ggaearintqwilltsgtteanawksltvghdfttkvpsaasidaakkagvnngnplday 182

QY 182 QQ 183
   ||
Db 183 qq 184

```

```

RESULT 10
ID P70492 standard; protein; 182 AA.
XX
AC P70492;
XX
DE 06-MAR-1991 (first entry)
XX
DE Streptavidin sequence.
XX
KW Streptavidin; N-terminal; fusion gene; fusion protein;
XX
OS Streptomyces avidin1.
XX
PN W08705026-A.
XX
PD 27-AUG-1987.

```

```

XX 24-FEB-1987; 87WO-US00397.
PF
XX 24-FEB-1986; 86US-0833324.
PR
XX (UYCO-) COLUMBIA UNIV N Y.
PA (UYNR-) UNIV NEW YORK.
XX
XX Cantor CR, Axel R, Garana C;
PI
XX WPI: 1987-250198/35.
DR N-PSDB; N70810.
XX
XX DNA encoding streptavidin - obtd. by restriction endo-nuclease
PT digestion of chromosomal DNA of Streptomyces avidin1
PT
XX
XX Disclosure; Figure 3; 54pp; English.
PS
XX
XX The sequence is that of streptavidin from S.avidin1. Streptavidin
CC may be expressed from a fusion gene comprising its coding gene and DNA
CC encoding a target protein of interest, where he streptavidin has
CC binding sites for biotin or deriv. The streptavidin has 4 binding
CC sites free for biotin, and is produced free of biotin
CC contamination. Improved streptavidins may also be produced by
CC site-directed mutagenesis. The fused gene may be used to produce
CC labeled, chemically-modified proteins in vivo, and to isolate
CC proteins when only the sequence of the gene is known.
XX
XX Sequence 182 AA:
SQ

```

```

Query Match          99.1%; Score 928; DB 8; Length 182;
Best Local Similarity 99.5%; Pred. No. 1.3e-75;
Matches 181; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 RKIVAAIAVSLTTVSTITSASADPSKSKAQYSAEAGITGTWYNQLGSTFIYTAGADG 61
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 rkivaaiaivslttvstitsasadpskdsraqysaaegltgwyngqlgstfiytagadg 60

QY 62 ALTGYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVAAMKNRYRNAHSATTMSGQYV 121
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 altgyesavгнаesryvltgrydsapadgsctalgwtvawkmnyrnahsattmsgyv 120

QY 122 GGAEARINTQWILTSGTTEANMWSKSLVGHDPFTTKVPSAASIDAAKKAGVNNGNPLDAY 181
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 ggaearintqwilltsgtteanawksltvghdfttkvpsaasidaakkagvnngnplday 180

QY 182 QQ 183
   ||
Db 181 qq 182

```

```

RESULT 11
ID Y84020 standard; protein; 183 AA.
XX
AC Y84020;
XX
DE 03-JUL-2000 (first entry)
XX
DE Amino acid sequence of the wildtype streptavidin monomer.
XX
KW Streptavidin; monomer; biotin binding domain; functional domain;
KW biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;
KW substrate; vascular device; prosthesis.
XX
OS Streptomyces avidin1.
XX
FH Key Location/Qualifiers
FT Peptide 1..24
FT /note= "propeptide"
XX
PN W0200011152-A1.

```

XX 02-MAR-2000.
 PD
 XX
 PF 25-AUG-1999; 99MO-US19481.
 XX
 PR 25-AUG-1998; 98US-0097816.
 XX
 PA (UNIW) UNIV WASHINGTON.
 PI Strayton PS, McDevitt TC, Nelson KE;
 DR WPI; 2000-224689/19.
 XX
 PT Mutant streptavidin molecule, useful as an adaptor and a coating for
 PT substrates such as vascular devices or prostheses, comprises a biotin
 PT binding domain and a secondary functional domain -
 PS Disclosure; Page 35-36; 43pp; English.
 XX
 CC The present sequence represents a wildtype streptavidin monomer. The
 CC specification describes streptavidin molecules comprising a biotin
 CC binding domain and a secondary functional domain. The molecules are
 CC adaptors with inherent effector function. They can therefore, bind to
 CC biotin and also have another function, for e.g. binding to a cell
 CC through the secondary functional domain comprising a cell adhesion
 CC peptide. The streptavidin molecules are useful as adaptors to bring,
 CC via a streptavidin/biotin interaction, the secondary functional domain
 CC into proximity with a cell or molecule to be affected and as a coating
 CC for substrates such as vascular devices or prostheses. Therefore, any
 CC compound of interest, such as a nucleic acid, protein, peptide, organic
 CC compound, inorganic compound, polysaccharide or a combination, can be
 CC targeted, delivered or immobilized using them. The molecules, when
 CC comprising an antibody for its secondary functional domain, are also
 CC useful in diagnostic applications for detecting analytes.
 XX
 XX Sequence 183 AA:
 SQ
 Query Match 99.1%; Score 928; DB 21; Length 183;
 Best Local Similarity 99.5%; Pred. No. 1.3e-75;
 Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MRKIVAAINSLTFTVITASASADPSKDSKAQVSAEAGITGWNOLGFTFVTAGAD 60
 Db 1 mrklvvaalavsltvtltsaasdpksdkagysaaegltgwnqlgscflvtagad 60
 QY 61 GALTGYESAVGNAESRYVLTGRYDAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120
 Db 61 galgtgyesavgnaesryvltgrydsapadgsgtalgwtvawkynyrmhsatcwsqgy 120
 QY 121 VGGAEARINTOMLTSSTTEANAMKSTLVGHDFRTKKYKPSAASIDAAKKAGVNNGNPLDA 180
 Db 121 vggaearinltqwtltsqgteanawksclvhdftrtkvkpsaasidaakkagvnngnplda 180
 QY 181 VQO 183
 Db 181 vqg 183
 RESULT 12
 ID W59217 standard; Protein: 183 AA.
 AC W59217;
 XX 27-AUG-1998 (first entry)
 XX
 DE S. avidinli streptavidin mutant protein #1.
 XX
 XX Streptavidin: ligand; binding affinity; mutant: isolation;
 KW purification; recover; immobilise.
 XX
 OS Streptomyces avidinli.

OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 68..71
 FT /label= VTAR68ESAV
 FT /note= "Wild type ESAV is replaced by VTAR. Numbering
 FT is from the start of the mature protein"
 XX
 XX EP835934-A2.
 XX
 PD 15-APR-1998.
 XX
 PF 09-OCT-1997; 97EP-0117504.
 XX
 PR 10-OCT-1996; 96DE-1041876.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 PI Skerra A, Voss S;
 DR WPI; 1998-218868/20.
 DR N-PSDB; V34715.
 XX
 XX This sequence represents a mutant streptavidin protein isolated from
 CC Streptomyces avidinli where the residues ESAV at position 44-47 of
 CC the mature wild type sequence are replaced by VTAR. This sequence is
 CC used to produce mutants which are used in a method to assay the binding
 CC affinity of streptavidin mutants. These mutants have a mutation within
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
 CC binding affinity than the wild-type for peptide ligands that include
 CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y
 CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant
 CC streptavidin mutants can be used to isolate, purify and determine
 CC proteins or to determine/recover substances that contain
 CC streptavidin-binding groups. Such compounds may also be used to
 CC immobilise fusions on microtitre plates, microbeads or sensor chips.
 CC NOTE: this sequence does not appear in the specification but has
 CC been constructed from the wild-type streptavidin sequence represented
 CC in V34714.
 XX
 XX Sequence 183 AA:
 SQ
 Query Match 98.2%; Score 919; DB 19; Length 183;
 Best Local Similarity 98.4%; Pred. No. 8.2e-75;
 Matches 180; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MRKIVAAINSLTFTVITASASADPSKDSKAQVSAEAGITGWNOLGFTFVTAGAD 60
 Db 1 mrklvvaalavsltvtltsaasdpksdkagysaaegltgwnqlgscflvtagad 60
 QY 61 GALTGYESAVGNAESRYVLTGRYDAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120
 Db 61 galgtgyvtaaragneesryvltgrydsapadgsgtalgwtvawkynyrmhsatcwsqgy 120
 QY 121 VGGAEARINTOMLTSSTTEANAMKSTLVGHDFRTKKYKPSAASIDAAKKAGVNNGNPLDA 180
 Db 121 vggaearinltqwtltsqgteanawksclvhdftrtkvkpsaasidaakkagvnngnplda 180
 QY 181 VQO 183
 Db 181 vqg 183
 RESULT 13
 ID W59218 standard; Protein: 183 AA.

```
XX AC W59218;
XX XX 27-AUG-1998 (first entry)
XX DE S. avidinii streptavidin mutant protein #2.
XX XX
XX XX Streptavidin; ligand; binding affinity; mutant; isolation;
XX KM purification; recover; immobilise.
XX OS Streptomyces avidinii.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 68..71 /label= IGAR68ESAV
XX FT /note= "Wild type ESAV is replaced by IGAR. Numbering
XX FT is from the start of the mature protein"
XX PN EP835934-A2.
XX PD 15-APR-1998.
XX XX
XX PF 09-OCT-1997; 97BP-0117504.
XX XX
XX PR 10-OCT-1996; 96DE-1041876.
XX XX
XX PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX PI Skerra A, Voss S;
XX XX
XX DR WPI; 1998-218868/20.
XX DR N-PSDB; V34716.
XX XX
XX PT Streptavidin mutants with higher binding affinity for peptide
XX PT ligands - have mutation in amino acid region 44-53, used to isolate,
XX PT purify or determine fusion proteins including these ligands
XX XX
XX PS Disclosure; Page -: 21pp; German.
XX XX
XX CC This sequence represents a mutant streptavidin protein isolated from
XX CC Streptomyces avidinii where the residues ESAV at position 44-47 of
XX CC the mature wild type sequence are replaced by IGAR. This sequence is
XX CC used to produce mutants which are used in a method to assay the binding
XX CC affinity of streptavidin mutants. These mutants have a mutation within
XX CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX CC binding affinity than the wild-type for peptide ligands that include
XX CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y
XX CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant
XX CC streptavidin mutants can be used to isolate, purify and determine
XX CC proteins or to determine/recover substances that contain
XX CC streptavidin-binding groups. Such compounds may also be used to
XX CC immobilise fusions on microtitre plates, microbeads or sensor chips.
XX CC NOTE: This sequence does not appear in the specification but has
XX CC been constructed from the wild-type streptavidin sequence represented
XX CC in V34714.
XX XX
XX SQ Sequence 183 AA;
```

```
Query Match 98.0%; Score 917; DB 19; Length 183;
Best Local Similarity 98.4%; Pred. No. 1.2e-74;
Matches 180; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 MKRIYVAALVSLTFTVSITASASADPSKDSKAQVSAEAGITGTWYNQIGSTFIWAGAD 60
DB 1 mrkivvaalavslftvtstasasadpskdskaqvsaeeagltgtwngstfiwagad 60
QY 61 GALTGTYESVAGNAESRYVLTGRYDSAPATDGGTALGWTVAAMKNRYRNAHSTTWSGOT 120
DB 61 galtgtiyagarnaesryvltgrysapatsdgsaltgwtvaknrynahnasattwsqgy 120
QY 121 VGGAEARINTQWMLTSGTTEANAMKSTLVGHDFETVKRPSAASIDAKKAGVNGNPLDA 180
```

```
DB 121 vggaearintqwltsgtteanawksclvyghdftkvkpsaasidaakkagvnmgnplda 180
QY 181 VQQ 183
DB 181 vqg 183
RESULT 14
Y84021
ID Y84021 standard; protein; 186 AA.
XX AC
XX XX Y84021;
XX DE 03-JUL-2000 (first entry)
XX XX
XX XX Amino acid sequence of a fibrinonectin/streptavidin fusion.
XX DE Streptavidin; monomer; biotin binding domain; functional domain;
XX KM biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;
XX KM substrate; vascular device; prosthesis; fibrinonectin.
XX OS Synthetic.
XX OS Streptomyces avidinii.
XX FH Key Location/Qualifiers
XX FT Peptide 1..24 /note= "propeptide"
XX FT
XX PN W020001152-A1.
XX XX
XX PD 02-MAR-2000.
XX XX
XX PF 25-AUG-1999; 99WO-US19481.
XX XX
XX PR 25-AUG-1998; 98US-0097816.
XX XX
XX PA (UNITW ) UNIV WASHINGTON.
XX PI Stayton PS, McDevitt TC, Nelson KE;
XX XX
XX DR WPI; 2000-224689/19.
XX XX
XX PT Mutant streptavidin molecule, useful as an adaptor and a coating for
XX PT substrates such as vascular devices or prostheses, comprises a biotin
XX PT binding domain and a secondary functional domain -
XX XX
XX PS Example 4; Page 36-37; 43pp; English.
XX XX
XX CC The present sequence represents a fusion of fibrinonectin and
XX CC streptavidin. It is a streptavidin molecule of the invention. The
XX CC specification describes streptavidin molecules comprising a biotin
XX CC binding domain and a secondary functional domain. The molecules are
XX CC adaptors with inherent effector function. They can therefore bind to
XX CC biotin and also have another function, for e.g. binding to a cell
XX CC through the secondary functional domain comprising a cell adhesion
XX CC peptide. The streptavidin molecules are useful as adaptors to bring,
XX CC via a streptavidin/biotin interaction, the secondary functional domain
XX CC into proximity with a cell or molecule to be affected and as a coating
XX CC for substrates such as vascular devices or prostheses. Therefore, any
XX CC compound of interest, such as a nucleic acid, protein, peptide, organic
XX CC compound, inorganic compound, polysaccharide or a combination, can be
XX CC targeted, delivered or immobilized using them. The molecules, when
XX CC comprising an antibody for its secondary functional domain, are also
XX CC useful in diagnostic applications for detecting analytes.
XX XX
XX SQ Sequence 186 AA;
```

```
Query Match 96.3%; Score 901.5; DB 21; Length 186;
Best Local Similarity 96.2%; Pred. No. 3e-73;
Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
```

```

QY 1 MKRIIVAAIAVSLTIVTSITASADSPSKDSKAQVSAEAAGITGTWYNQLGSTFIYTAGAD 60
  |||
DB 1 mrklvvaalavsltlvtvsltasadpskdskaqvsaaagltglwvnyqlgstflvttagad 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPA---TDGSGTALGWTAVWKNKNYRNASHATTWS 117
  |||
DB 61 galtgtyesavgnaesryvltgrysdapgtdpgsgtalgwtvawkmyrnashattws 120
QY 118 GQYVGAAGARINTQWLTSCTTEANAMKSTLVGHDFTFKVPSAASIDAAKKAGVNNGNP 177
  |||
DB 121 gqyvgaagearintqwltsqteanawkscllvehdftfkvpsaasidaakkagvnngrp 180
QY 178 LDAVQQ 183
  |||
DB 181 ldavq 186

RESULT 15
Y84022
Y84022 standard; protein: 186 AA.
XX
AC Y84022;
XX
DT 03-JUL-2000 (first entry)
XX
DE Amino acid sequence of a osteopontin/streptavidin fusion.
XX
KW Streptavidin; monomer; biotin binding domain; functional domain;
KW biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;
KW substrate; vascular device; prosthesis; osteopontin.
XX
OS Synthetic.
OS Streptomyces avidin11.
FH Key Location/Qualifiers
FT 1..24
FT Peptide /note= "propeptide"
PN WO200011152-A1.
XX
PD 02-MAR-2000.
XX
PF 25-AUG-1999; 99WO-US19481.
XX
PR 25-AUG-1998; 98US-0097816.
XX
PA (UNITV ) UNITV WASHINGTON.
PI Stayton PS, McDevitt TC, Nelson KE;
XX
DR WPI; 2000-224689/19.
XX
PT Mutant streptavidin molecule, useful as an adaptor and a coating for
PT substrates such as vascular devices or prostheses, comprises a biotin
PT binding domain and a secondary functional domain -
XX
PS Disclosure; Page 38; 43pp; English.
XX
CC The present sequence represents a fusion of osteopontin and
CC streptavidin. It is a streptavidin molecule of the invention. The
CC specification describes streptavidin molecules comprising a biotin
CC binding domain and a secondary functional domain. The molecules are
CC adaptors with inherent effector function. They can therefore, bind to
CC biotin and also have another function, for e.g. binding to a cell
CC through the secondary functional domain comprising a cell adhesion
CC peptide. The streptavidin molecules are useful as adaptors to bring,
CC via a streptavidin/biotin interaction, the secondary functional domain
CC into proximity with a cell or molecule to be affected and as a coating
CC for substrates such as vascular devices or prostheses. Therefore, any
CC compound of interest, such as a nucleic acid, protein, peptide, organic
CC compound, inorganic compound, polysaccharide or a combination, can be
CC targeted, delivered or immobilized using them. The molecules, when
CC comprising an antibody for its secondary functional domain, are also

```

```

CC useful in diagnostic applications for detecting analytes.
XX
SQ Sequence 186 AA;
  Query Match 96.1%; Score 899.5; DB 21; Length 186;
  Best Local Similarity 96.2%; Pred. No. 4,6e-73;
  Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
QY 1 MKRIIVAAIAVSLTIVTSITASADSPSKDSKAQVSAEAAGITGTWYNQLGSTFIYTAGAD 60
  |||
DB 1 mrklvvaalavsltlvtvsltasadpskdskaqvsaaagltglwvnyqlgstflvttagad 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPA---TDGSGTALGWTAVWKNKNYRNASHATTWS 117
  |||
DB 61 galtgtyesavgnaesryvltgrysdapgtdpgsgtalgwtvawkmyrnashattws 120
QY 118 GQYVGAAGARINTQWLTSCTTEANAMKSTLVGHDFTFKVPSAASIDAAKKAGVNNGNP 177
  |||
DB 121 gqyvgaagearintqwltsqteanawkscllvehdftfkvpsaasidaakkagvnngrp 180
QY 178 LDAVQQ 183
  |||
DB 181 ldavq 186

```

Search completed: April 6, 2001, 00:15:59
 Job time: 15499 sec


```
|||||
250 CGAGTCGCGCCGTCGCAACGCCGAGAGCCGCTACGCTCGACCGGTCTT 299
84 YRASPSEALAPROALATHRAPGLYSERGLYTHRALALEUGLYTRPHR 100
300 ACGACAGGCCCCGGCCACCGACGCGACCGCCCTCGGTGGAGC 349
101 VALALATPLYSASNANTYRARGASNALAHISSEALATHRTPRSE 117
350 GTGGCTCGAAGANTAACTACCGCAACCCCACTCCGGACCACTGAG 399
117 TGLYLNTYVALGLYALAGLUALAARGILEASNTHRGINTPLEUL 134
400 CGGCACGACGTCGCGCGCGCGAGCGAGATCAACACCCAGTGGCTGC 449
134 EUTHRSERGLYTHRTHRGUALASNAATPLYSERTHLEUVALGLY 150
450 TGACCTCGGACCAACCGACGCGCTGGAAGTCCACGCTGGTCGCG 499
151 HISAPTHRPHERTHLYSVALYSPROSERALALASERILEASPALAAL 167
500 CACGACACCTTCACCAAGGTGAAGCCGCGCCCTCCATCGACGCGC 549
167 ALYLSYALAGLYVALASNSNGLYASNPROLEUASPALAVAGLNGIN 183
550 GAAGAAGCGCGCGCTCAACAACGCAACCCGCTCGACGCGCTTCAGCAG 598
```

seq_name: gb_pat1:A93649

```
seq_documentation_block: 638 bp DNA PAT 22-JAN-2000
LOCUS A93649
DEFINITION Sequence 1 from Patent EP0799890.
ACCESSION A93649
VERSION A93649.1 GI:6741838
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 638)
AUTHORS Mueller,R.D. and Deger,A.D.
TITLE Recombinant inactive core streptavidin mutants
JOURNAL Patent: EP 0799890-A 08-OCT-1997;
BOEHRINGER MANNHEIM GMBH (DE)
FEATURES
source 1..638
sig_peptide /organism="unidentified"
CDS sig_peptide /db_xref="taxon:32644"
50..121
122..598 /note="unnamed protein product"
/codon_start=1
/protein_id="CAB69587.1"
/db_xref="GI:6741839"
/translation="MKIVVAIAVSLTTSITASADPSKDSKAQVSAABAGITGT
WYNQGSTPIVYAGADALGTYESAVGNAESRYVLTGRYDAPADTDSGTALGWTV
WNNYRNHSAATWMSGOYVGAEARINTQMLLTSGTTEANMKNSTLVGHDTFTKXPS
ASISDAKKAGANNONPLDAVQO"
```

```
mat_peptide 122..598
BASE COUNT 115 a 244 c 193 g 86 t
ORIGIN
```

alignment_scores:

Quality: 936.00 Length: 183
Ratio: 5.115 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-589-870-2 x A93649 ..

Align seq 1/1 to: A93649 from: 1 to: 638

1 MejArlgylsilevalalalalalalalalserleuthrthrvalse 17

```
|||||
50 ATGCGCAAGATCGTCGTTGACCATGCGCCTTCCCTGACCAAGGCTCTC 99
17 TLETHRALASERIALASERIALASPPROSERLYSASPSEALAGLVN 34
100 GATTACGGCCAGCCTTCGCGACACCCCTCCAAAGACTCGAAGGCCAGG 149
34 ALSERIALAGLUALAGLYLETHRGLYTRTPYRASNGLNGLY 50
150 TCTGCGCGCGGAGCGCGCATCACCGGACCTGATACCACTCGGC 199
51 SERTHRPHLEVALTRHALAGLYALASPGLYALAEUTHRGINTYR 67
200 TCGACCTTCATCGTAGCGCGCGCGAGCGCGCCCTCGACGGAACTA 249
67 TGLUSERALAVAGLYVALASNALGUSERATGYRVALLEUTHRGILYRGT 84
250 CGAGTCGCGCCGTCGCAACGCCGAGAGCCGCTACGCTCGACCGGTCTT 299
84 YRASPSEALAPROALATHRAPGLYSERGLYTHRALALEUGLYTRPHR 100
300 ACGACAGGCCCCGGCCACCGACGCGACGCGCCTCGGTGGAGC 349
101 VALALATPLYSASNANTYRARGASNALAHISSEALATHRTPRSE 117
350 GTGGCTCGAAGANTAACTACCGCAACCCCACTCCGGACCACTGAG 399
117 TGLYLNTYVALGLYALAGLUALAARGILEASNTHRGINTPLEUL 134
400 CGGCACGACGTCGCGCGCGCGAGCGAGATCAACACCCAGTGGCTGC 449
134 EUTHRSERGLYTHRTHRGUALASNAATPLYSERTHLEUVALGLY 150
450 TGACCTCGGACCAACCGACGCGCTGGAAGTCCACGCTGGTCGCG 499
151 HISAPTHRPHERTHLYSVALYSPROSERALALASERILEASPALAAL 167
500 CACGACACCTTCACCAAGGTGAAGCCGCGCCCTCCATCGACGCGGC 549
167 ALYLSYALAGLYVALASNSNGLYASNPROLEUASPALAVAGLNGIN 183
550 GAAGAAGCGCGCGCTCAACAACGCAACCCGCTCGACGCGCTTCAGCAG 598
```

seq_name: gb_pat1:101349

```
seq_documentation_block: 638 bp ss-DNA PAT 21-MAY-1993
LOCUS 101349
DEFINITION Sequence 2 from Patent US 4839293.
ACCESSION 101349
VERSION 101349.1 GI:270135
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 638)
AUTHORS Cantor,C.R., Axel,R. and Argarana,C.
TITLE DNA encoding streptavidin, streptavidin produced therefrom, fused
polypeptides which include amino acid sequences present in
streptavidin and uses thereof
JOURNAL Patent: US 4839293-A 2 13-JUN-1989;
The Trustees of Columbia University in the City of New York;
New York, NY
```

```
FEATURES
source 1..638
BASE COUNT 114 a 244 c 194 g 86 t
ORIGIN
```

alignment_scores: 933.00 Length: 183
Ratio: 5.098 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.454

Alignment block:

US-09-589-870-2 x I01349 ..

Align seg 1/1 to: I01349 from: 1 to: 638

```

1 MetArgLysIleValAlaIleAlaIleAlaValSerLeuThrThrValSe 17
  |||||||
50 ATGGCAAGATCGTCGTTGCAGCATCGCCCTTCCCGACGACGCGTCTC 99
  |||||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
  |||||||
100 GATTACGGCCGACGCTTCGCGACACCCCTCCAGAGACTCGAGGCGCAGG 149
  |||||||
34 aIleSerAlaAlaGlnAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
  |||||||
150 TCTCGCGCGCGGAGCGCGCATCACCGGCACCTGTACACACCTCTCGGC 199
  |||||||
51 SerThrPheIleValThrAlaGlyAlaAlaAspGlyAlaLeuThrGlyThrTy 67
  |||||||
200 TCGACCTTCATCGTGACCGCGCGCGCGACGCGCCCTGACCGGAACTTA 249
  |||||||
67 rGluSerAlaValAlaGlyAlaSerArgTyrValLeuThrGlyArgT 84
  |||||||
250 CGAGTCGGCGCGTCGCAACGCGCGAGCGCGCTACGTCCTGACCGGTGCTT 299
  |||||||
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
  |||||||
300 ACGACAGCGCGCCGCGCCACGACGCGAGCGCGCGCCCTGCTGGGAGC 349
  |||||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
  |||||||
350 GTGGCTCGGAAGAACTAACTACCGCAACGCCCACTCCGCGACGCTGGAG 399
  |||||||
117 rGlyGlnTyrValAlaGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuL 134
  |||||||
400 CGGCGAGTACGTCGCGCGCGCGCGAGGAGATCAACACCCAGTGGCTGC 449
  |||||||
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
  |||||||
450 TGACCTCGGCGACCAACGAGGCAACGCTGGAAGTCCACGCTGCTCGGC 499
  |||||||
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
  |||||||
500 CACGACACCTTCACCAAGGTGAAGCGCTCGCGCCCTCATCGACGCGGC 549
  |||||||
167 aLysLysAlaGlyAlaValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
  |||||||
550 GAAGAAGCGCGCGCTCAACACAGGCAACCGCGCTGACGCGCTTCAGCAG 598
  |||||||

```

seq_name: gb_ba3:S78777

seq_documentation_block:

LOCUS S78777 625 bp DNA BCT 30-OCT-1995
 DEFINITION streptavidin v1 (Streptomyces violaceus, Tu 2460, Genomic, 625 nt).
 ACCESSION S78777
 VERSION S78777.1 GI:1042193
 KEYWORDS
 SOURCE Streptomyces violaceus Tu 2460.
 ORGANISM Streptomyces violaceus
 Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
 Actinomycetales: Streptomyces: Streptomyces.

REFERENCE 1 (bases 1 to 625)

AUTHORS Bayer, E.A., Kulik, T., Adar, R. and Wilchek, M.

TITLE Close similarity among streptavidin-like, biotin-binding proteins from Streptomyces

JOURNAL Biochim. Biophys. Acta 1263 (1), 60-66 (1995)

MEDLINE Genbank staff at the National Library of Medicine created this entry [NCBI glibsg 169185] from the original journal article.

REMARK This sequence comes from Fig. 4.

FEATURES

1..625

Location/Qualifiers

source

```

/organism="Streptomyces violaceus"
/db_xref="taxon:1936"
50..601
/gene="streptavidin v1, Sa v1"
50..601
/gene="streptavidin v1, Sa v1"
/note="This sequence comes from Fig. 5; Sa v1"
/product="streptavidin v1"
/protein_id="AB35015.1"
/db_xref="GI:1042194"
/translation="MKRIIVAAIYAVSLTFTVSTIASASADPSKSKAQAQVAEAGITGT
WYNQGSFTFAGADGALGTGESAVENSRVYLTGRYDSA PATDGSSTALGWTVA
WKNVYRNAHSATFTWSGOYGVGTGARINTQMLTLSTGTENAMKSTLVGHDTFTKVRPS
MSIDAKKAGVNNCPNDAVD0"
BASE COUNT 115 a 240 c 186 g 84 t
ORIGIN

```

```

alignment_scores:
  Quality: 932.00 Length: 183
  Ratio: 5.093 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.454

```

alignment_block:

US-09-589-870-2 x S78777 ..

Align seg 1/1 to: S78777 from: 1 to: 625

```

1 MetArgLysIleValAlaIleAlaIleAlaValSerLeuThrThrValSe 17
  |||||||
50 ATGGCAAGATCGTCGTTGCAGCATCGCCCTTCCCGACGACGCGTCTC 99
  |||||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
  |||||||
100 GATTACCGCCGACGCTTCGCGACACCCCTCCAGAGACTCGAGGCGCAGG 149
  |||||||
100 GATTACCGCCGACGCTTCGCGACACCCCTCCAGAGACTCGAGGCGCAGG 149
  |||||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
  |||||||
100 GATTACCGCCGACGCTTCGCGACACCCCTCCAGAGACTCGAGGCGCAGG 149
  |||||||
34 aIleSerAlaAlaGlnAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
  |||||||
150 TCTCGCGCGCGGAGCGCGCATCACCGGCACCTGTGTACACACCTCTCGGC 199
  |||||||
51 SerThrPheIleValThrAlaGlyAlaAlaAspGlyAlaLeuThrGlyThrTy 67
  |||||||
200 TCGACCTTCATCGTGACCGCGCGCGCGACGCGCCCTGACCGGAACTTA 249
  |||||||
67 rGluSerAlaValAlaGlyAlaSerArgTyrValLeuThrGlyArgT 84
  |||||||
250 CGAGTCGGCGCGTCGCAACGCGCGAGCGCGCTACGTCCTGACCGGTGCTT 299
  |||||||
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
  |||||||
300 ACGACAGCGCGCGCGCGCATCACCGGCACGCGCGCGCTGCTGGGAGC 349
  |||||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
  |||||||
350 GTGGCTCGGAAGAACTAACTACCGCAACGCCCACTCCGCGACGCTGGAG 399
  |||||||
117 rGlyGlnTyrValAlaGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuL 134
  |||||||
400 CGGCGAGTACGTCGCGCGCGCGCGAGGAGATCAACACCCAGTGGCTGC 449
  |||||||
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
  |||||||
450 TGACCTCGGCGACCAACGAGGCAACGCTGGAAGTCCACGCTGCTCGGC 499
  |||||||
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
  |||||||
500 CACGACACCTTCACCAAGGTGAAGCGCTCGCGCCCTCATCGACGCGGC 549
  |||||||
167 aLysLysAlaGlyAlaValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
  |||||||
550 GAAGAAGCGCGCGCTCAACACAGGCAACCGCGCTGACGCGCTTCAGCAG 598
  |||||||

```

```

seq_name: gb_ba3:S78782
seq_documentation_block:
LOCUS       S78782             625 bp    DNA             BCT       30-OCT-1995
DEFINITION   streptavidin v2 [Streptomyces violaceus, Tu 2605, Genomlc, 625 nt].
ACCESSION    S78782
VERSION      S78782.1 GI:1042195
KEYWORDS
SOURCE       Streptomyces violaceus Tu 2605.
ORGANISM     Streptomyces violaceus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.
REFERENCE    1 (bases 1 to 625)
AUTHORS      Bayer, E.A., Kulik, T., Adar, R. and Wilchek, M.
TITLE        Close similarity among streptavidin-like, biotin-binding proteins
            from Streptomyces
JOURNAL      Biochim. Biophys. Acta 1263 (1), 60-66 (1995)
MEDLINE      93359204
REMARK       Genbank staff at the National Library of Medicine created this
            entry [NCBI gidsq 169186] from the original journal article.
            This sequence comes from Fig. 4.
FEATURES
    source
        location/Qualifiers
            1..625
                /organism="Streptomyces violaceus"
                /db_xref="taxon:1936"
                /gene="streptavidin v2, Sa v2"
                /note="This sequence comes from Fig. 5; Sa v2"
                /codon_start=1
                /product="streptavidin v2"
                /protein_id="AAB35016.1"
                /db_xref="GI:1042196"
                /translation="MKYIVAAIVASLTGVGTASASADPSKSKQAQVAENGITGT
                WYNQAGSTIVANADGSLITGTESAVAESRYIVTGDAPADDSGTLAWGTVA
                WNNIRNASATVNSGOYVAGSEARINTQMLTSGTTAANAMKSTLVGHDTEFKVPS
                AASIDAARAGAVNNGNPLDAYOQ"
BASE COUNT   117 a      244 c      184 g      80 t
ORIGIN
alignment_scores:
    quality: 897.00      Length: 183
    Ratio: 4.929        Gaps: 0
    Percent Similarity: 99.454      Percent Identity: 95.082
alignment_block:
    US-09-589-870-2 x S78782 ..
Align seg 1/1 to: S78782 from: 1 to: 625
1 MetArgLysIleValAlaAlaAlaIleAlaValSerIleuThrValSe 17
|||||
50 ATGCGCAAGATCGTGTGACGACCATTTGCTGACACACCGTCGG 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGly 34
|||||
100 TATATAGCGCAGCGCCCTCCGGGAGCCCGTCGAAGACTCCAGGCCAGG 149
|||||
34 aISerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
|||||
150 CCGCGCTTCGCGGAGCGCGCATCCCGCACCTGCTCAACACGACTCGGT 199
|||||
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyTrpTyr 67
|||||
200 TCGACCTTCAATCGACCGCAAGCCGACGACGACCTCAACCGGCACCTA 249
|||||
67 rGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValIleuThrGlyArgT 84
|||||
250 CGAGTCGGCGGTGCGCAACCGCAAGACCGCTACGCTCAACCGGTGTT 299
|||||
84 YrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100

```

```

seq_name: gb_pat1:A39565
seq_documentation_block:
LOCUS       A39565             1356 bp    DNA             PAT       05-MAR-1997
DEFINITION   Sequence 4 from Patent WO9415644.
ACCESSION    A39565
VERSION      A39565.1 GI:2295847
KEYWORDS
SOURCE       unidentified.
ORGANISM     unidentified.
REFERENCE    1 (bases 1 to 1356)
AUTHORS      Epenetos, A.A., Spooner, R.A. and Deonrain, M.
TITLE        COMPOUNDS FOR TARGETING
JOURNAL      Patent: WO 9415644-A 4 21-JUL-1994;
            IMP CANCER RES TECH (GB)
            Other publication GB 2289679 951129.
FEATURES
    source
        location/Qualifiers
            1..1356
                /organism="unidentified"
                /db_xref="taxon:32644"
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAA02449.1"
                /db_xref="GI:2295848"
                /translation="MKYLLPTAAAGLLILAOPAMAQVOYLOQPGAELVYPGASVYLSG
                KASGYPTSYNMMHWKQRPGRLEWIGRIDPNSGCTKYNEKFKSKATITVDPKPSSTAY
                MQLSLTSEDSKAVYICARFDYRGSIFDYWCQTTVYSSGGGSGGGSGGGGQAV
                VTQESALTTSPGETVTLTCRSTGAVTSSNANWQEPDHLFTGLIGTNNRAAGVP
                ARFSGSLIGDKRALITGAQTEDEAIYFCALMYSNHWVFGGCTKLTIVGLEPAAPAP
                DPSKDSKAQVSAEAGITGTWYNQGSFFIYTAGDGLATGTVESAAGVNASRYVLTG
                RYDSAPATDGSSTGLATGWAMKNNRNANASFTWSGOYVGAEEARINTQMLTSGTTE
                ANAMKSTLVGHDTEFKVPSAASIDAARAGAVNNGNPLDAYOQ"
BASE COUNT   332 a      357 c      361 g      306 t
ORIGIN
alignment_scores:
    quality: 846.00      Length: 164
    Ratio: 5.190        Gaps: 0
    Percent Similarity: 99.390      Percent Identity: 98.780
alignment_block:
    US-09-589-870-2 x A39565 ..
Align seg 1/1 to: A39565 from: 1 to: 1356
20 AlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAl 36
|||||

```

```
853 GCTGCCGACCTGCAGACCCGTCACAGACTCCAAAGCTCAGTTTCTGC 902
36 aalagluualaglylthrrgylthrrprrasglnleuglyserthr 53
|||||
903 AGCCGAAGCTGTATCACTGCGACCTGTATTAACCACTGGGGTCGACT 952
53 hellervalthralaglyalaaspalyalaueuthrglythrtyrgluser 69
|||||
953 TCATTGTGACCGCTGCTGCGACGAGCTCTGACTGGCAGCTACGAACTCT 1002
70 AlavAlglYasnaAlgluserArgrYvalleuthrglyArgrYraSpse 86
|||||
1003 GCGGTTGGTAAACGAGAAATCCCGCTACGTACTGACGCGCTATGACTC 1052
86 rAlaPrAlaThrAspGlySerGlyThrAlaLeuGlyTrpThrValAlaT 103
|||||
1053 TGCACCTGCCACCGATGCTGTGTACCGCTGCGCTGCGACTGTGCTT 1102
103 rPlYsAsnAsnTyrArGAsnAlaHsSerAlaThrTrpSerGlyGln 119
|||||
1103 GGAACAAACMACTATCTATATCCGACAGCCGCTACGTGCTGCCCAA 1152
120 TyrValGlyglYalagluAlaArgIleasnThrGlnTrpLeuLeuThrSe 136
|||||
1153 TACGTTGGCGGCTGCTGAGGCTCGTATCAACACTCAGTGGCTGTTAACTC 1202
136 rGlyThrThrGluAlaAsnAlaTrpGlySerThrLeuValGlyHisAspT 153
|||||
1203 CGGCACTACCGAAGCGAATGCATGAAATCGACACTAGTAGTCATGACA 1252
153 hrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLys 169
|||||
1253 CCTTTACCAAAAGTTAAAGCTTCTGCTGTACATTGATGCTGCCAAGAA 1302
170 AlaGlyValAsnaSnglyAsnProLeuAspAlaValGlnGln 183
|||||
1303 GCAGCGCGTAAACACGGAACCTCTAGACGCTGTTCAGCAA 1344
seq_name: gb_pat1:A93152
seq_documentation_block:
LOCUS A93152 1356 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 4 from Patent EP0815872.
ACCESSION A93152
VERSION A93152.1 GI:6741540
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos,A.A. and Deonarain,M.
TITLE Compounds for targeting
JOURNAL Patent: EP 0815872-A 07-JAN-1998;
IMP CANCER RES TECH (GB)
FEATURES
source
1..1356
location/Qualifiers
CDS
1..1347
/db_xref="taxon:32644"
/note="unnamed protein product"
/protein_id="CAB69540.1"
/db_xref="GI:6741541"
/translacion="MKYLLPTAAAGLLLLAOPAMAQVQLQPGAEIVKPGASVKLS
KASGYFTSYMMHWKQRPGRGLEWIGRIDPNSGGRKYNEKRSKATLTVDPSSTAY
NOLSYLTSDSAVYVCARDYGSYEDWGQTYTVSSGGSGSGSGSGSGSGAV
VHOESALTSPGETVITICRSSTGAVTSTNANWQEKPDHLEFTGLIGTNNRAPGP
AARPSGLIDKRAALITTCQTEDRAITFPALVSNHWYGGCKTLVLGLIENAPAA
DPSKDSKAQVSAAEAGITGTWYNOLGSTIVTAGADGALGTYESAVNAESRYVLTG
RVDSAFATDGSALGTVAMKNNRNNAHSATTWGQVYGAEARINTWMLTSGTTE
AAMKSTLGHDTFTVKRPSAASIDAAKAGVNGNPDAVQ0"
```

```
alignment_scores:
Quality: 846.00 Length: 164
Ratio: 5.190 Gaps: 0
Percent Similarity: 99.390 Percent Identity: 98.780
alignment_block:
US-09-589-870-2 x A93152
Align seg 1/1 to: A93152 from: 1 to: 1356
20 AlAserAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAl 36
|||||
853 GCTGCCGACCTGCAGACCCGTCACAGACTCCAAAGCTCAGTTTCTGC 902
36 aalagluualaglylthrrgylthrrprrasglnleuglyserthr 53
|||||
903 AGCCGAAGCTGTATCACTGCGACCTGTATTAACCACTGGGGTCGACT 952
53 hellervalthralaglyalaaspalyalaueuthrglythrtyrgluser 69
|||||
953 TCATTGTGACCGCTGCTGCGACGAGCTCTGACTGGCAGCTACGAACTCT 1002
70 AlavAlglYasnaAlgluserArgrYvalleuthrglyArgrYraSpse 86
|||||
1003 GCGGTTGGTAAACGAGAAATCCCGCTACGTACTGACGCGCTTATGACTC 1052
86 rAlaPrAlaThrAspGlySerGlyThrAlaLeuGlyTrpThrValAlaT 103
|||||
1053 TGCACCTGCCACCGATGCTGTGTACCGCTGCGCTGCGACTGTGCTT 1102
103 rPlYsAsnAsnTyrArGAsnAlaHsSerAlaThrTrpSerGlyGln 119
|||||
1103 GGAACAAACMACTATCTATATCCGACAGCCGCTACGTGCTGCCCAA 1152
120 TyrValGlyglYalagluAlaArgIleasnThrGlnTrpLeuLeuThrSe 136
|||||
1153 TACGTTGGCGGCTGCTGAGGCTCGTATCAACACTCAGTGGCTGTTAACTC 1202
136 rGlyThrThrGluAlaAsnAlaTrpGlySerThrLeuValGlyHisAspT 153
|||||
1203 CGGCACTACCGAAGCGAATGCATGAAATCGACACTAGTAGTCATGACA 1252
153 hrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLys 169
|||||
1253 CCTTTACCAAAAGTTAAAGCTTCTGCTGTACATTGATGCTGCCAAGAA 1302
170 AlaGlyValAsnaSnglyAsnProLeuAspAlaValGlnGln 183
|||||
1303 GCAGCGCGTAAACACGGAACCTCTAGACGCTGTTCAGCAA 1344
seq_name: gb_pat1:AR082490
seq_documentation_block:
LOCUS AR082490 1356 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 4 from patent US 5973116.
ACCESSION AR082490
VERSION AR082490.1 GI:10009216
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos,A.Antoniou, Spooner,R,Anthony and Deonarain,M.
TITLE Compounds for targeting
JOURNAL Patent: US 5973116-A 4 26-OCT-1999;
FEATURES
source
1..1356
location/Qualifiers
CDS
1..1356
/db_xref="taxon:32644"
/note="unnamed protein product"
/protein_id="CAB69540.1"
/db_xref="GI:6741541"
/translacion="MKYLLPTAAAGLLLLAOPAMAQVQLQPGAEIVKPGASVKLS
KASGYFTSYMMHWKQRPGRGLEWIGRIDPNSGGRKYNEKRSKATLTVDPSSTAY
NOLSYLTSDSAVYVCARDYGSYEDWGQTYTVSSGGSGSGSGSGSGSGAV
VHOESALTSPGETVITICRSSTGAVTSTNANWQEKPDHLEFTGLIGTNNRAPGP
AARPSGLIDKRAALITTCQTEDRAITFPALVSNHWYGGCKTLVLGLIENAPAA
DPSKDSKAQVSAAEAGITGTWYNOLGSTIVTAGADGALGTYESAVNAESRYVLTG
RVDSAFATDGSALGTVAMKNNRNNAHSATTWGQVYGAEARINTWMLTSGTTE
AAMKSTLGHDTFTVKRPSAASIDAAKAGVNGNPDAVQ0"
```


JOURNAL Patent: WO 9728263-A 07-AUG-1997;
 LUBITZ WERNER (AT); SLEVTR UWE (AT)
 FEATURES Location/Qualifiers
 source 1..498
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 120 a 137 c 131 g 110 t
 ORIGIN

alignment_scores:
 Quality: 834.00 Length: 159
 Ratio: 5.245 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-589-870-2 x A93838 ..

Align seg 1/1 to: A93838 from: 1 to: 498

```

25 AspproserlysaSpserlysaIaGlnValSerAlaIaGluIaGlyI 41
   |||||||
7  GACCCGTCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGAAAGCTGAT 56
41 eThrGlyThrTPrTyrsngInLeuGlySerThrPheIleValThrIaG 58
   |||||||
57 CACTGGCACTGCTGATATACCAACTGGGGTCGACTTTCATTGTGAACGCTG 106
58 IyAlaAspGlyAlaLeuThrGlyThrGlySerAlaValGlyAla 74
   |||||||
107 GTGGGGAGGAGCTGCTGACTGCACCTACGAATCTGCGTTGTGAACGCA 156
75 GluserArgTyValLeuThrGlyArgTyAspSerAlaProAlaThrAs 91
   |||||||
157 GAATCCCGCTACGACTGACTGGCCGTTATGACTGTGCACCTGCCACCGA 206
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyra 108
   |||||||
207 TGGCTGTGGTACCGCTCTGGGCTGAGCTGGCTTGGAANAACAATATC 256
108 rGaAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyValGlyAla 124
   |||||||
257 GTAATGGGCACAGCGCCACTACGTGCTGGCCAATACGTTGGCGGCT 306
125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
   |||||||
307 GAGCGCTGTATCAACACTGAGTGGCTGTTAACATCCGCACTACCGAAGC 356
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158
   |||||||
357 GAATGCATGGAAATCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 406
158 ySProSerAlaIaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
   |||||||
407 AGCCTTCTGCTGCTAGCATGATGCTGCCAAGAAACAGCGCGTAAACAC 456
175 GlyAsnProLeuAspAlaValGlnGln 183
457 GGTAAACCTCTAGACGCTGTTCAGCAA 483
seq_name: gb_pat1:A00743
seq_documentation_block:
LOCUS A00743 507 bp DNA PAT 28-JAN-1993
DEFINITION S.avidinii synthetic gene for streptavidin.
ACCESSION A00743
VERSION A00743.1 GI:14605
KEYWORDS
SOURCE
ORGANISM Streptomyces avidinii.
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE "1 (bases 1 to 507)

```

AUTHORS Edwards, R.
 TITLE SYNTHETIC GENE
 JOURNAL Patent: WO 8903422-A 2 20-APR-1989;
 British Bio-Technology Ltd
 FEATURES Location/Qualifiers
 source 1..507
 /organism="Streptomyces avidinii"
 /db_xref="taxon:1895"
 10..492
 /note="synthetic gene"
 /codon_start=1
 /transl_table=11
 /product="streptavidin"
 /protein_id="CAA00084.1"
 /db_xref="GI:14606"
 /translation="MDPSKDSKAQVSAARAGITGTYNOIGSTFTVTAGACALIGTY
 ESAVGVASRSRYVLTKRYDSAPATDSDGTALGVWAKKNYRAHSATTWGQVSGAE
 ARINTQMLTSGTTEANAMKSTLVGHDPTRKVPAAASIDAAKAGVNNGNPLDAVQO
 "

BASE COUNT 125 a 137 c 130 g 115 t
 ORIGIN

alignment_scores:
 Quality: 834.00 Length: 159
 Ratio: 5.245 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-589-870-2 x A00743 ..

Align seg 1/1 to: A00743 from: 1 to: 507

```

25 AspproserlysaSpserlysaIaGlnValSerAlaIaGluIaGlyI 41
   |||||||
13  GACCCGTCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGAAAGCTGAT 62
41 eThrGlyThrTPrTyrsngInLeuGlySerThrPheIleValThrIaG 58
   |||||||
63 CACTGGCACTGCTGATATACCAACTGGGGTCGACTTTCATTGTGAACGCTG 112
58 IyAlaAspGlyAlaLeuThrGlyThrGlySerAlaValGlyAla 74
   |||||||
113 GTGGGAGCGAGCTCTGACTGGCACTACGAATCTGGGTTGTGAACGA 162
75 GluserArgTyValLeuThrGlyArgTyAspSerAlaProAlaThrAs 91
   |||||||
163 GAATCCCGCTACGACTGACTGGCCGTTATGACTGTGCACCTGCCACCGA 212
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyra 108
   |||||||
213 TGGCTGTGGTACCGCTCTGGGCTGAGCTGGCTTGGAANAACAATATC 262
108 rGaAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyValGlyAla 124
   |||||||
263 GTAATGGGCACAGCGCCACTACGTCGCTGCCAATACGTTGGCGGCT 312
125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
   |||||||
313 GAGCGCTGTATCAACACTGAGTGGCTGTTAACATCCGCACTACCGAAGC 362
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158
   |||||||
363 GAATGCATGGAAATCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 412
158 ySProSerAlaIaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
   |||||||
413 AGCCTTCTGCTGCTAGCATGATGCTGCCAAGAAACAGCGCGTAAACAC 462
175 GlyAsnProLeuAspAlaValGlnGln 183
463 GGTAAACCTCTAGACGCTGTTCAGCAA 489

```



```
seq_name: gb_pat1:A00744
seq_documentation_block: 507 bp DNA PAT 28-JAN-1993
LOCUS A00744
DEFINITION S.avidinii synthetic gene (reverse complement) for streptavidin.
ACCESSION A00744
VERSION A00744.1 GI:14607
KEYWORDS
SOURCE
ORGANISM Streptomyces avidinii.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS Edwards, R.
TITLE SYNTHETIC GENE
JOURNAL Patent: WO 8903422-A 3 20-APR-1989;
British Bio-Technology Ltd
COMMENT #A00744 is the reverse complement of #A00743.
FEATURES
source 1..507
/organism="Streptomyces avidinii"
/db_xref="taxon:1895"

BASE COUNT 115 a 130 c 137 g 125 t
ORIGIN

alignment_scores:
Quality: 834.00 Length: 159
Ratio: 5.245 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x A00744/rev ..

Align seg 1/1 to reverse of: A00744 from: 1 to: 507

25 ASPROSERLYASPSERLYSALAGINVALSERIALAAGLUALAGLYI 41
|||||
495 GACCGGTCCAGGACCTCCAAAGCTCAGTTCTGCAGCCGGAAGCTGGTAT 446
|||||
41 ethgLyThrPTyRAsngInLeuGlySerThrPheIleValThrAlaG 58
|||||
445 CACTGGCACCTGATATACCAACTGGGGTGCACATTTCATGTGACCGCTG 396
|||||
58 LyaIAspGlyAlaLeuThrGlyThrTyrgIuSerAlaValGlyAsnAla 74
|||||
395 GTGGGAGGAGGAGCTCTGACTGGCACTACGAATCTGGCGTTGGTAAACCA 346
|||||
75 GluSerArgTyValIleuThrGlyArgTyRAspSerAlaProIatThrAs 91
|||||
345 GAATCCCGCTACGTACTGAGCGCGTTATGACTCTGCACCTGCCACCGCA 296
|||||
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyra 108
|||||
295 TGGCTGTGTACCGCTGTGGCTGGAGCTGGCTTGAAGAAAACAACATATC 246
|||||
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyValGlyGlyAla 124
|||||
245 GTAATGCCACAGCGCCACTAGTGGCTGGCCAATATACGTTGGCGGTCT 196
|||||
125 GluAlaArgGlyLeuAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
|||||
195 GAGGCTGCTATCAACACTCAGTGGCTGTTAACAATCCGGCACTACCGAAGC 146
|||||
141 AAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
|||||
145 GAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAAAGTTA 96
|||||
158 ySPROSERAlaIAserTleAspAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
95 AGCCTTCTGCTGCTAGCAATTGATGCTGCCAAGAAAGCAGGCGTAAACAAC 46
|||||
175 GlyAsnProLeuAspAlaValGlnGln 183
```

```
|||||
45 GGTAACCTCTAGACGCTGTACAGCA 19
|||||

seq_name: gb_pat1:A39567
seq_documentation_block: 1296 bp DNA PAT 05-MAR-1997
LOCUS A39567
DEFINITION Sequence 6 from Patent WO9415644.
ACCESSION A39567
VERSION A39567.1 GI:2295849
KEYWORDS
SOURCE
ORGANISM unidentified.
unidentified.
unclassified.

REFERENCE
AUTHORS Epenetos, A.A., Spooner, R.A. and Deonaraia, M.
TITLE COMPOUNDS FOR TARGETING
JOURNAL Patent: WO 9415644-A 6 21-JUL-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication GB 2289679 951129.
FEATURES
source 1..1296
/organism="unidentified"
/db_xref="taxon:32644"
40..1287
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02450.1"
/db_xref="GI:2295850"
/translation="MKYLLPTAAAGLLLLAQPAMAOVOLQOPGAEIVKPGASVYKLSK
KASGYFTSYMMHWVKORPGRLEWIGRIDPNSGTKYNEKFSKATLYVDKPSSTAY
MOLSLTSEDSAVYVCARYDYGSYEDYWGQTTVTYSSGGSGSGSGSGSGSGSAG
VNOESALTISRGETVTLTCRSSGATYTSNNAMVOKEPDHLFTGLIGTNNAPAPV
ARFSGSLIDKNAALITTGQTEDEALITFCALMTSNHWVGGTKLVGLIEAPAAPA
DPSKDSKAOVSAAEAGITGTWYNQISTEIVTAGADGALVTYESAVNAESRYVLTG
RYDSAPATDGSSTALGMIVAMKNNYNRNASHATTWSGOYVGAEARINTOMLTSCTTE
ANAMKSTLVGHDFYTKVKSAS"

BASE COUNT 312 a 342 c 347 g 295 t
ORIGIN

alignment_scores:
Quality: 745.00 Length: 144
Ratio: 5.210 Gaps: 0
Percent Similarity: 99.306 Percent Identity: 98.611

alignment_block:
US-09-589-870-2 x A39567 ..

Align seg 1/1 to: A39567 from: 1 to: 1296

20 AlAserAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAl 36
|||||
853 GCTCCGCGACCTGCAGACCCGTCCAAGGAGCTCCAAAGCTCAGGTTTCTGC 902
|||||
36 aAlaGlnAlaGlyIleThrGlyThrTrpTyRAsngInLeuGlySerThrP 53
|||||
903 AGCCGAACCTGTATCACTGGCACCTGGTATPACCACTGGAGGCTGACTT 952
|||||
53 heIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrgIuSer 69
|||||
953 TCATTGTACCGCTGCTGCGAGCGAGCTGTGACGTGCACTACCAATCT 1002
|||||
70 AlAvalGlyAsnAlaGluSerArgTyValIleuThrGlyArgTyRAspSe 86
|||||
1003 GCGGTGGTAAACGAGAAATCCCGTACCTACTGACTGCGCGCTTATGACTC 1052
|||||
86 rAlaProIatThrAspLysSerGlyThrAlaLeuGlyTrpThrValAlaTr 103
|||||
1053 TGCACCTGCACCGATGGCTGTGTTACCGCTGGCGTGGAGCTGGGCTT 1102
|||||
103 rPLysAsnAsnTyRArgAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
|||||
```

1103 GGAACCACTATCGTATCGGCACAGCCACTACGTGGTGGCCAA 1152
120 TyrValGlyGlyAlaGluAlaArgIleasnThrGlnTrpLeuThrSe 136
|||||
1153 TACGTGGCGGCTGCTGAGGCTCGTATCAACAACAGTGGCTGTACATC 1202
136 rGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspT 153
|||||
1203 CGGCACCTACCGACGAGCAATGCATGGAATCGACACTAGTACGTCATGACA 1252
153 hrPheThrLysValLysProSerAlaAlaSer 163
|||||
1253 CCTTACCAGTTAAGCCTTCTGCTGCTAGC 1284


```

1 ATGCGCAGATCGTGTGACGCCATCGCCGTTTCCCTGACGACGGTCTC 50
PT r1leThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGln 34
51 GATTACGGCCAGCCCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAGG 100
34 a1SerAlaAlaGlnAlaGly1leThrGlyThrTrpTyrAsnGlnLeuGly 50
101 TCTCGCGCGCCGAGCGCGCATCACCGGACCTGGTACCAACACACTCGGC 150
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThr 67
151 TCGACCTTCATCGTACCGCGCGCGCGACGCGCCCTGACCGGACCTTA 200
67 rGluSerAlaValGlyAlaGlnAlaGlnSerArgTyrValLeuThrGlyArg 84
201 CGAATCGGCGCTCGCAACGCGGAGAGCGGTACGCTCGACCGGTCGTT 250
84 y1AspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
251 ACGACAGCGCCCGCCACCGACGACGCGGACCGCCCTGGTTGGAGC 300
101 ValAlaThrPylsAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSe 117
301 GTGGCTGGAGGAATTAATACCGCAACGCCACCTCCGCGACACGTGGAG 350
117 rGlyGlnTyrValGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 134
351 CGGCACTACGTCGCGCGCGCGCGGCGGAGGATTAACACCCAGTGGCTGC 400
134 eutHrSerGlyThrThrGlnAlaAsnAlaThrPylsSerThrLeuValGly 150
401 TGACCTCGCGACCAACGAGGCCAAGCGCTGGAGTCCACGCTGGTGGCC 450
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAla 167
451 CACGACACCTTCACCAAGGTGAAGCGCTCGCGCTCCATCGACGCGGC 500
167 alySLyAlaGlyAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
501 GAACAAGCGCGCGCTCAACAAGCGCAACCGCTCGACGCGCTTCAGCAG 549
seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:291073
seq_documentation_block:
ID 291073 standard; DNA; 563 BP.
XX
AC 291073;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomycetes avidinii sps gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; sps gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomycetes avidinii.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX
PK 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
WPI: 2000-195402/17.
DR P-PSDB: Y80512.

```

```

XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 84; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomycetes
CC avidinii sps gene as an example of the heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 563 BP; 107 A; 213 C; 169 G; 74 T; 0 other:

```

```

alignment_scores:
    Quality: 936.00      Length: 183
    Ratio: 5.115        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

```

```

alignment_block:
US-09-589-870-2 x 291073

```

```

Align seg 1/1 to: 291073 from: 1 to: 563

```

```

1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrValSe 17
10 ATGGCGCAAGATCGCTGTCGACGATCGCGCTTCCCTGACCAACGGCTC 59
17 r1leThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGln 34
60 GATTACGGCCAGCGCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAGG 109
34 a1SerAlaAlaGlnAlaGly1leThrGlyThrTrpTyrAsnGlnLeuGly 50
110 TCTCGCGCGCGGAGCGCGGCATCACCGGACCTGGTACCAACGACTCGGC 159
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThr 67
160 TCGACCTTCATCGTACCGCGCGCGCGGCGCGCTCGACCGGAACCTTA 209
67 rGluSerAlaValGlyAlaGlnAlaGlnSerArgTyrValLeuThrGlyArg 84
210 CGAATCGGCGCTCGCAACGCGGAGAGCGGTACGCTCGACCGGTCGTT 259
84 y1AspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
260 ACGACAGCGCCCGCGCACCGGACGCGGCAACCGCCCTGGTTGGAGC 309
101 ValAlaThrPylsAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSe 117
310 GTGGCGTGGAGGAATTAATACCGCAACGCCACCTCCGCGACACGTGGAG 359
117 rGlyGlnTyrValGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 134
360 CGGCACTACGTCGCGCGCGCGGAGGAGATCAACCCAGTGGCTGC 409
134 eutHrSerGlyThrThrGlnAlaAsnAlaThrPylsSerThrLeuValGly 150
410 TGACCTCGCGACCAACGAGGCCAAGCTGGAAGTCCAGCTGGTGGCG 459
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAla 167
460 CACGACACCTTCACCAAGGTGAAGCGCTCGCGCTCCATCGACGCGGC 509

```

167 alvLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
510 GAGAGAGCGCGCGTCACACACGCAACCGCCGTCGACGCCGTCGACGAG 558

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT:T73193

seq_documentation_block:

ID T73193 standard; DNA; 638 BP.

XX T73193;

XX 27-APR-1998 (first entry)

XX wild-type streptavidin DNA.

DE Streptavidin: biotin; anti-interference reagent; detection: mutetin;
KM avidin: non-specific binding; ss.

XX Unidentified.

XX Key Location/Qualifiers

FT CDS 50..601

FT /tag- a

FT /product- streptavidin

FT sig_peptide 50..121

FT /tag- b

FT mat_peptide 122..598

FT /tag- c

FT /product- streptavidin

XX DE19637718-A1.

XX 02-OCT-1997.

XX 16-SEP-1996; 96DE-1037718.

XX 01-APR-1996; 96DE-1013053.

XX (BOE) BOEHRINGER MANNHEIM GMBH.

XX Brandstetter H, Deger A, Engh R, Kopetzki E, Mueller R;

XX Schmitt U;

XX WPI; 1997-482043/45.

XX P-PSDB; W29306.

XX Streptavidin and avidin muteins with reduced binding affinity for
PT biotin - useful for reducing interference from nonspecific binding
PT in assays

XX Disclosure; Page 16-17; 26pp; German.

XX This DNA sequence encodes a streptavidin which is used in a novel method
CC of reducing interference from non-specific binding in assays. Muteins
CC constructed from a core streptavidin or avidin sequence are selected that
CC differ from the native polypeptide by at least one amino acid and have a
CC binding affinity for biotin of less than 1010 l/mole. The biotin-bindable
CC polypeptide may be present as a polymeric conjugate, e.g. with another
CC polypeptide or protein, especially bovine serum albumin. These muteins
CC are used as anti-interference reagents for reducing and/or avoiding
CC nonspecific interactions in a process for detecting an analyte. In
CC particular, they are used in assays where the streptavidin/avidin-biotin
CC specific binding pair is involved for qualitative and/or quantitative
CC determination of an analyte in a test sample, e.g. a heterogeneous
CC immunoassay or a hybridisation assay. Despite having a lower binding
CC affinity for biotin, the muteins have high immunological cross-reactivity
CC with native streptavidin and avidin.

XX Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

XX Alignment_scores:

Quality: 936.00 Length: 183
Ratio: 5.115 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-589-870-2 x T73193 ..

Align seg 1/1 to: T73193 from: 1 to: 638

1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSe 17
|||||
50 ATGCGCAAGATCGCTTGCAGACCATCGCGTTCCTCCGACGACGCTCTC 99
17 rleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34
|||||
100 GATTACGGCCAGCGCTTGCGCAGACCTCCCAAGAGACTGAAAGGCCAGG 149
34 alSerAlaAlaGlyAlaGlyIleThrGlyThrTrpTyraSnglnLeuGly 50
|||||
150 TCTCGCGCCCGAGCGCGCATACCGGCACCTGCTGATACACAGCTCGGC 199
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
200 TCGACCTTCATCTGTGACCGCGCGCGCCGACGCGCCCTGACCGGAACCTA 249
67 rGlSerAlaValAlaGlyAsnAlaGlyIleSerArgTyraValLeuThrGlyArgT 84
|||||
250 CGAGTCGGCGCTCGGCAAGCGCGAGCGCGCTACCTCTGACCGGCTCTT 299
84 yraSPSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyraThr 100
|||||
300 ACGACAGCGCGCCCGCGCCACGACGCGACGCGCCCTGCTGGTTGGACG 349
101 ValAlaTrpLysAsnAsnTyraArgAsnAlaHisSerAlaThrThrTrpSe 117
|||||
350 GTGGCTGTGAAGAAATTACTACCGCAACGCCCATCTCCGCGACACGCTGGAG 399
117 rGlyGlnTrpValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrpLeuL 134
|||||
400 CGGCGCTGACTCGCGCGCGCGCGAGGATCAACACCGACGCTGCG 449
134 euThrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCGGCGCACCGACGCGCAAGCGCTGGAAGTCCACGCTGGTGGC 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
|||||
500 CACGACACCTTCACCAAGGTGAAGCGCTCGCGCTCCATCGACGCGCG 549
167 alvLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 GAGAGAGCGCGCGTCACACACGCAACCGCCGTCGACCGCTTCGACGAG 598

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT:V34714

seq_documentation_block:

ID V34714 standard; DNA; 638 BP.

XX V34714;

XX 27-AUG-1998 (first entry)

XX S. avidinii streptavidin cDNA.

DE Streptavidin: ligand; binding affinity: mutant; isolation;

KM purification: recover; immobilise; ss.

XX Streptomycetes avidinii.

XX OS

XX Key Location/Qualifiers

FT CDS 50..601

FT /tag- a

FT sig_peptide 50..121

FT mat_peptide /tag= b
 FT 122..598
 FT /tag= c
 FT /product= Streptavidin

EP835934-A2.

15-APR-1998.

09-OCT-1997; 97EP-0117504.

10-OCT-1996; 96DE-1041876.

(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

Skerra A, Voss S;

WPI; 1998-218868/20.

P-PSDB; W59216.

Streptavidin mutants with higher binding affinity for peptide ligands - have mutation in amino acid region 44-53, used to isolate, purify or determine fusion proteins including these ligands

Disclosure; Page : 21pp; German.

This sequence encodes a wild-type streptavidin protein isolated from Streptomyces avidinii. This sequence is used to produce mutants which are used in a method to assay the binding affinity of streptavidin mutants. These mutants have a mutation within the amino acid (aa) region 44-53 of the wild-type protein show a higher binding affinity than the wild-type for peptide ligands that include the sequence of formula

TRP-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips.

NOTE: This sequence does not appear in the specification but is used to make the mutant streptavidin sequence represented in V34715 and V34716.

Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

alignment_scores:
 Quality: 936.00 Length: 183
 Ratio: 5.115 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-589-870-2 x V34714 ..

Align seg 1/1 to: V34714 from: 1 to: 638

1 MetArgLysIleValIleAlaIleAlaValSerLeuThrValse 17
 |||||
 50 ATGGCAAGATCGTCGTGACGATCGCGTTCCTGACACGAGCTC 99
 |||||
 17 rlieThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34
 |||||
 100 GATTACGGCGCGCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAGG 149
 |||||
 34 alserAlaIaIaIaIaIaIaIleThrGlyThrTrpTyrAsnGlnLeuGly 50
 |||||
 150 TCTGGCGCGCGGAGCGCGGATCAGCGGACCTGGTACAAACCACTCGGC 199
 |||||
 51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
 |||||
 200 TCGACCTTCATCGTACCGGCGCGGCGGACGCGCTCGACGGAACCTA 249
 |||||
 67 rGIuSerAlaValaIaIaValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 84
 |||||
 250 CGAGTCGGCGCGTGGCAACGCCGAGAGCGCTAGCTCCTGACCGCGTGT 299

seq_name: /cgcn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:249867

seq_documentation_block:
 ID 249867 standard; cDNA; 638 BP.

AC 249867;

DT 25-APR-2000 (first entry)

DE Streptavidin gene for recombinant pARR27 vector.

XX Potato proteinase inhibitor-II; PPI-II; streptavidin; worm;

XX insect; plant-noxious protein; pest resistance; moth; insect; weevil;

XX grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;

XX insecticidal; ss.

XX Synthetic.

XX	OS	Key	Location/Qualifiers
XX	FT	CDS	50..601
XX	FT	sig_peptide	50..121
XX	FT	mat_peptide	122..598
XX	FT		/tag= b
XX	FT		/tag= c

WO200004049-A1.

PD 27-JAN-2000.

XX 15-JUL-1999; 99WO-NZ00110.

XX 15-JUL-1998; 98NZ-0331002.

XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.

XX Christeller JT, Sutherland PW, Murray C, Markwick NP, Philip BA;

XX Malone LA, Burgess EPU;

XX WPI; 2000-171244/15.

XX P-PSDB; Y44701.

XX New chimeric polypeptide and composition comprising the polypeptide

XX useful for conferring pest resistance on plants

XX Disclosure; Fig 12; 11pp; English.

530 GATTACGGCCAGCGCTCGCGACACCCCTCCAGGACTCGAAGGCCAGG 579

34 a1SerAla1aGlu1aGly1IleThrGly1ThrTrp1Asn1Gln1Leu1Gly 50
 580 TCTCGCGCGCCGAGGCGGATCACCAGCCTGATACAACTGAGCTCGGC 629

51 Ser1rhe1IleVal1ThrAla1Gly1Ala1Asp1Gly1Ala1Leu1Thr1Gly1 67
 630 TCGACCTTCATCGTGACCGCGCGGCCGACGCGCCCTGACCGGAACTA 679

67 rGluSer1aVal1aGly1aAsn1aGlu1Ser1aArg1aTyr1aVal1euthr1aGly1a 84
 680 CGAGTCGGCCCGCTCGGCAACGCCGAGAGCCGCTACGCTCGACCGGTCCTT 729

84 YrAspSer1aPro1aThr1aAsp1aGly1aSer1aGly1aThr1aLeu1aGly1aThr1a 100
 730 ACGACAGCGCCCGCGCCACCGAGCGAGCGGACCGCCCTCGGTTGGACG 779

101 Val1a1aTrp1aLys1aAsn1aSer1aArg1aAsn1aAla1aHis1aSer1aThr1aTrp1a 117
 780 GTGGCCTGGAAGAAATTAATACCGCAAGCCCACTCGCGACAGCTGGAG 829

117 rGly1aIn1aTyr1aVal1aGly1aAla1aArg1aIle1aAsn1aThr1aGln1aLeu1a 134
 830 CGGCCAGTACGTCGCGCGCGCGGAGGAGATCAACACCCAGTGGCTGC 879

134 euthrSer1aGly1aThr1aGlu1aAla1aSer1aTrp1aLys1aSer1aThr1aLeu1aGly1a 150
 880 TGACCTCGCGGACCAACCGAGGCCCAACGCTTGGAACTCCACGCTGGTCGC 929

151 His1aSP1rhe1rhe1rLys1aVal1aLys1aPro1aSer1aAla1aSer1aLeu1aAla1a 167
 930 CACGACACCTTCACCAAGGTGAAGCCGTCGCGCGCTCCATCGAAGCGGC 979

167 aLys1aLys1aGly1aVal1aAsn1aGly1aAsn1aPro1aLeu1aAsp1aVal1aGln1a 183
 980 GAAGAAGGCGCGCGCTCAACAAGCGCAACCCGCTCGACCGCTTCAGCAG 1028

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:291074

seq_documentation_block:
 ID 291074 standard; DNA; 564 BP.

XX 291074;
 XX
 XX 06-JUN-2000 (first entry)
 XX
 XX Streptomyces avidinii mst gene.
 XX
 XX Streptomyces avidinii mst gene.
 XX
 XX Plant somatic tissue degeneration; plant essential factor; depletion;
 XX viability; mst gene; plant development; plant morphology; flower;
 XX fruit plant; ds.
 XX
 XX Streptomyces avidinii.
 OS
 XX WO200007427-A2.
 PN
 XX 17-FEB-2000.
 PD
 XX 30-JUL-1999; 99WO-IL00420.
 PF
 XX 03-AUG-1998; 98IL-0125632.
 PR
 XX (AGRI-) AGRIC RES ORG.
 XX
 XX PI Kapulink Y, Ginzberg I;
 XX
 XX WPI: 2000-195402/17.
 DR P-PSDB; Y80513.
 XX
 XX Degeneration of somatic plant tissue by expression of a heterologous
 PT protein, useful for controlling plant development and morphology, such
 PT as decreasing the number of flowers present to increase the number of

PT fruit -
 XX
 XX Examples: Page 85; 91pp; English.
 PS
 XX The invention relates to a method of effecting degeneration of a somatic
 CC plant tissue by expressing a heterologous protein capable of binding a
 CC plant essential factor (PEF), in somatic plant tissue cells, where
 CC heterologous protein expression causes depletion of the PEF so the plant
 CC viability is maintained, while simultaneous degeneration of the somatic
 CC plant tissue is effected. This sequence represents the Streptomyces
 CC avidinii mst gene as an example of a heterologous gene introduced into
 CC the plants. The methods can provide for the selective and optionally
 CC reversible cell degeneration in somatic plant tissue. They can be used
 CC for artificially controlling plant development and morphology. They can
 CC be used e.g. to decrease the number of flowers in fruit producing plants
 CC so as to increase the number of fruits which reach maturity.
 XX
 S0 Sequence 564 BP; 109 A; 213 C; 169 G; 73 T; 0 other;

alignment_scores:
 Quality: 931.00 Length: 182
 Ratio: 5.115 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-589-870-2 x 291074 ..

Align seg 1/1 to: 291074 from: 1 to: 564

2 Arg1aLys1aVal1aAla1aIle1aVal1aSer1aLeu1aThr1aVal1aSer1a 18
 14 CGCAAGATCGTCGTTGACGACCATGCGCTTCCCTGACCAAGCTCTGCAT 63
 18 eThr1aLys1aSer1aAla1aAsp1aPro1aLys1aAsp1aSer1aLys1aAla1aVal1a 35
 64 TACGGCAGCGCTTCGCGAGACCCCTCCCAAGAGCTCAAGGCCAGGCTCT 113
 35 eAla1aAla1aGlu1aGly1aIle1aThr1aGly1aThr1aTrp1aAsn1aGln1aLeu1aSer 51
 114 CGCGCGCGGAGCGCGGATCAACCGGACCTGTACAAACAGCTCGGCTCG 163
 52 Thr1rhe1IleVal1ThrAla1aGly1Ala1aAsp1aGly1Ala1aLeu1aThr1aTrp1a 68
 164 ACCTTCATCGTGACCGCGCGGCCGCGAGCGGCCCTTGACCGGAACTACGA 213
 68 uSer1aVal1aGly1aAsn1aGlu1aSer1aArg1aTyr1aVal1euthr1aGly1aArg1aTyr 85
 214 GTGCGCGCTCGGCAACCGCGAGCGCGCTACGTCCTGACCGGTCGTACG 263
 85 sPse1aAla1aPro1aThr1aAsp1aGly1aSer1aGly1aThr1aAla1aLeu1aGly1aTrp1a 101
 264 ACAAGCGCGCGCGGACCGAGCGAGCGGACCGCCCTCGGCTGGAGCGTG 313
 102 Ala1aTrp1aLys1aAsn1aTrp1aArg1aAsn1aAla1aHis1aSer1aAla1aThr1aTrp1a 118
 314 GCTGGAAGAAATTAATACCGCAAGCCCACTCCGCAACGAGTGAACGG 363
 118 yGln1aTyr1aVal1aGly1aAla1aArg1aIle1aAsn1aThr1aGln1aTrp1aLeu1a 135
 364 CCAGTACGTCGCGCGCGCGGAGGAGATCAACACCAAGTGGCTGTGA 413
 135 hTrser1aGly1aThr1aGlu1aAla1aTrp1aLys1aSer1aThr1aLeu1aGly1aHis 151
 414 CCTCGGCAACCGAGGCGCAACGCGCTGGAATGCAACGCTGTGGCGCAC 463
 152 Asp1rhe1rhe1rLys1aVal1aLys1aPro1aSer1aAla1aSer1aLeu1aAla1aVal1a 168
 464 GACACCTTCACCAAGGTGAAGCCGTCGCGCTCCATCGACCGGCGGAA 513
 168 sLys1aAla1aGly1aVal1aAsn1aGly1aAsn1aPro1aLeu1aAsp1aVal1aGln1a 183
 514 GAAGCGCGCGGCTCAACAAGCGCAACCGGCTCGAGCGGCTTCACACAG 559

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA1993.DAT:Q53412

seq_documentation_block:

ID Q53412 standard; DNA; 552 BP.

AC Q53412;

DT 27-JUN-1994 (first entry)

DE Streptavidin gene.

KM Streptavidin; protein secretion; Bacillus subtilis; ss.

OS Streptomyces avidin11.

XX Key Location/Qualifiers

FT CDS 1..552

FT mat_peptide /*tag= a

FT sig_peptide /*tag= b

FT misc_feature /*tag= c

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

FT /note= "fused to npr signal peptide in pBE673"

34 a1SerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50

101 TCTCGCGCGCGAGCGCGGCATCGCGCACTGTCACACACCTCGCGC 150

51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67

151 TCGACCTTCATCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 200

67 rGluSerAlaValGlyAsnAlaGluSerArgTyrValLeuThrGlyThrTy 84

201 CGAGTCGCTT 250

84 YrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100

251 ACGACACGCTT 300

101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrpPse 117

301 GTGCGCTGGAAGTAATTAACCGCACCGCGCGCGCGCGCGCGCGCGCGCTT 350

117 rGlyGlnTyrValGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134

351 CGGCGAGTACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 400

134 eutThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150

401 TGACCTCGCTT 450

151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167

451 CACGACACCTTCACCAAGGTGAAGCGTCGCGCGCGCGCGCGCGCGCGCGCTT 500

167 alyLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183

501 GAAGAAGCTT 549

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA1998.DAT:V34715

seq_documentation_block:

ID V34715 standard; DNA; 638 BP.

AC V34715;

DT 27-AUG-1998 (first entry)

DE S. avidin11 streptavidin mutant #1 DNA.

KM Streptavidin; ligand; binding affinity; mutant; isolation;

KW purification; recover; immobilise; ss.

XX Streptomyces avidin11.

OS Synthetic.

XX Key

FT CDS 50..601

FT sig_peptide /*tag= a

FT mat_peptide /*tag= b

FT /product= Streptavidin

FT mutation 251..262

FT /tag= d

FT /note= "Wild type GAGTCGCGCGCTC is replaced by

FT GTCACGCGCGCT"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 PA
 XX
 PI Skerra A, Voss S;
 XX
 DR WPI: 1998-218868/20.
 DR P-PSDB; W59217.
 XX
 PT Streptavidin mutants with higher binding affinity for peptide
 PT ligands - have mutation in amino acid region 44-53, used to isolate,
 PT purify or determine fusion proteins including these ligands
 XX
 PS
 XX
 XX
 CC This sequence encodes a mutant streptavidin protein isolated from
 CC Streptomyces avidinii where the residues ESAV at position 44-47 of
 CC the mature wild type sequence are replaced by VVAR. This sequence is
 CC used to produce mutants which are replaced in a method to assay the binding
 CC affinity of streptavidin mutants. These mutants have a mutation within
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
 CC binding affinity than the wild-type for peptide ligands that include
 CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y
 CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant
 CC streptavidin mutants can be used to isolate, purify and determine
 CC proteins or to determine/recover substances that contain
 CC streptavidin-binding groups. Such compounds may also be used to
 CC immobilise fusions on microtitre plates, microbeads or sensor chips.
 CC NOTE: This sequence does not appear in the specification but has
 CC been constructed from the wild-type streptavidin sequence represented
 CC in V34714.
 XX
 SQ Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

alignment_scores:
 Quality: 919.00 Length: 183
 Ratio: 5.077 Gaps: 0
 Percent Similarity: 98.907 Percent Identity: 98.361

alignment_block:

US-09-589-870-2 x V34715 ..

Align seg 1/1 to: V34715 from: 1 to: 638

1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSe 17
 |||||
 50 ATGCGCAGAGTCGTGGACCCATCGCCGTCCTCGACACGAGTCCTC 99
 17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34
 |||||
 100 GATTACGGCCAGCGCTTCGGCAGACCCCTCCAAAGACTCCGAAGCCCAAG 149
 34 aISerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
 |||||
 150 TCTCGGCGCGCGAGCGCGCATCACCGGCACTGGTACAACAGACTCGGC 199
 51 SerThrPheIleValThrAlaGlyAlaAspGlyValAlaLeuThrGlyTyr 67
 |||||
 200 TCGACCTTCATCTGGACCGCGGCGCGGACGCGCCCTCGACCGGAACCTTA 249
 67 rGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
 |||||
 250 CGTCACCGCGCGGTGGCAACCGCGAGACCGCTACGTCTACCGGTCGT 299
 84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPThr 100
 |||||
 300 ACACACACCGCGCGCGCACCGACGCGACGCGACCGCCCTCGGTGGACG 349
 101 ValAlaIleThrLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
 |||||
 350 GTGGCCCTGGAGAAATACTAACCGCAAGCCCACTCCGCGACCAACGGTAG 399
 117 rGlyGlnTyrValGlyValAlaGluAlaArgIleAsnThrGlnTrpLeuL 134

|||||
 400 CGGCCAGTACTCGCGCGCGCGAGCGAGATCAACACCCAGTGCTGC 449
 134 eunhrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
 |||||
 450 TGACCTCGGCAACCGAGCGCAACGCGCTGAGATCGTCACGCTGTCGGC 499
 151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
 |||||
 500 CACGACACCTTCACCAAGGTGAAAGCCGTCGCCGCTTCATCGACGCGGC 549
 167 aLysAlaAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
 |||||
 550 GAACAGAGCCGCGCCCTCAACAGCAACGCCCTCGACGCGCTTACGACG 598
 seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT:V34716
 seq_documentation_block:
 ID V34716 standard; DNA: 638 BP.
 XX
 AC V34716;
 XX
 DT 27-AUG-1998 (first entry)
 XX
 DE S. avidinii streptavidin mutant #2 DNA.
 XX
 KW Streptavidin; ligand; binding affinity; mutant; isolation;
 KW purification; recover; immobilise; ss.
 XX
 OS Streptomyces avidinii.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 50..601
 FT /tag= a
 FT sig_peptide 50..121
 FT /tag= b
 FT mat_peptide 122..598
 FT /tag= c
 FT /product= Streptavidin
 FT mutation 251..262
 FT /tag= d
 FT /note= "Wild type GAGTCGCGCTC is replaced by
 ATCGTCGAGC"
 FT
 XX
 PN EP835934-A2.
 XX
 PD 15-APR-1998.
 XX
 PF 09-OCT-1997; 97EP-0117504.
 XX
 PR 10-OCT-1996; 96DE-1041876.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 XX
 PI Skerra A, Voss S;
 XX
 DR WPI: 1998-218868/20.
 DR P-PSDB; W59218.
 XX
 PT Streptavidin mutants with higher binding affinity for peptide
 PT ligands - have mutation in amino acid region 44-53, used to isolate,
 PT purify or determine fusion proteins including these ligands
 XX
 PS
 XX
 XX
 CC This sequence encodes a mutant streptavidin protein isolated from
 CC Streptomyces avidinii where the residues ESAV at position 44-47 of
 CC the mature wild type sequence are replaced by IGAR. This sequence is
 CC used to produce mutants which are replaced in a method to assay the binding
 CC affinity of streptavidin mutants. These mutants have a mutation within
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
 CC binding affinity than the wild-type for peptide ligands that include

CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y
 CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant
 CC streptavidin mutants can be used to isolate, purify and determine
 CC proteins or to determine/recover substances that contain
 CC streptavidin-binding groups. Such compounds may also be used to
 CC immobilise fusions on microtitre plates, microbeads or sensor chips.
 CC NOTE: This sequence does not appear in the specification but has
 CC been constructed from the wild-type streptavidin sequence represented
 CC in V34714.

XX
 XX Sequence 638 BP; 116 A; 242 C; 194 G; 86 T; 0 other;

alignment_scores:

Quality: 917.00 Length: 183
 Ratio: 5.066 Gaps: 0
 Percent Similarity: 98.907 Percent Identity: 98.361

alignment_block:

US-09-589-870-2 x V34716 ..

Align seg 1/1 to: V34716 from: 1 to: 638

```

1 MeArghYsIleValAlaAlaIleAlaValSerLeuThrThrValSe 17
|||||
50 ATGGCGAAGATCGTGTGACGACCATGCGCTTCCCTGACGACGCTCTC 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34
|||||
100 GATTACCGCCGACCTTGGCAGACCCCTCCAAAGAGACTCGAAGGCCAGG 149
|||||
34 AlSerAlaAlaGlnAlaGlyIleThrGlyThrPTyrAsnGlnLeuGly 50
|||||
150 TCTCGGCGCGGAGCGCGCATCACCGGACCTGTACTCAACACCTCGGC 199
|||||
51 SerThrHeIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
200 TCGACCTTCATCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 249
|||||
67 rGluSerAlaValAlaGlyIleThrGlyThrPTyrValLeuThrGlyArgT 84
|||||
250 CATGCGCTGGAGGAATTAACCGCAAGCCGCTGCTGCTGCTGCTGCTT 299
|||||
84 YrAspSerAlaProAlaThrAspGlyThrPTyrAlaLeuGlyThrPTyr 100
|||||
300 ACGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 349
|||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrpPse 117
|||||
350 GTGGCTCGAAGGAATTAACCGCAAGCCGCTGCTGCTGCTGCTGAG 399
|||||
117 rGlyGlnTyrValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuL 134
|||||
400 CGGCGACAGACTCGCGCGCGCGCGCGAGGAGATCAACACCCAGTGGCTGC 449
|||||
134 eutThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCTCGGACACCGAGGCGCAACGCTGGAGATCCACACCTGCTGCGC 499
|||||
151 HisAspThrPheThrLysValLysProSerAlaIleAspSerIleAspAla 167
|||||
500 CACGACACCTTCACCAAGGTGAAGCCGCTCGCGCGCTTCATCGACGCGGC 549
|||||
167 dLysLysAlaGlyValAlaAsnGlnLysProLeuAspAlaValGlnGln 183
|||||
550 GAAGAAGCGCGCGCTCAACAACGCAACCGCGCTGACGCGCTTGACGAG 598
|||||

```

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q70659

seq_documentation_block:

ID Q70659 standard; DNA; 1356 BP.

XX
 AC Q70659;

XX
 DT 26-MAR-1995 (first entry)

XX
 DE ScFv PRAS108 and PRAS112 DNA.

XX
 KW Amplification; single chain variable region fusion protein; PCR; ss.

XX
 OS Synthetic.

XX
 FH Key Location/Qualifiers

FT CDS 40..1348

FT /tag= a

XX
 PN MO9415644-A.

XX
 PD 21-JUL-1994.

XX
 PF 17-JAN-1994; 94MO-GB00087.

XX
 PR 15-JAN-1993; 93GB-0000686.

XX
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

XX
 PI Deonarain M, Epenetos AA, Spooner RA;

XX
 DR WPI: 1994-248907/30.

XX
 DR P-PSDB; R56483.

XX
 PT New cpds. comprising a targeting portion and a cytotoxic portion
 PT - used esp. for treating mammals for destroying target cells,
 PT partic. tumour cells

XX
 PS Claim 36; Fig 9; 114pp; English.

XX
 CC The sequence is that of the ScFv PRAS108 and PRAS112 between HindIII and
 CC EcoRI site obtd. by PCR.

XX
 CC See also Q70652-69.

XX
 SQ Sequence 1356 BP; 332 A; 359 C; 359 G; 306 T; 0 other;

alignment_scores:

Quality: 846.00 Length: 164
 Ratio: 5.190 Gaps: 0
 Percent Similarity: 99.390 Percent Identity: 98.780

alignment_block:

US-09-589-870-2 x Q70659 ..

Align seg 1/1 to: Q70659 from: 1 to: 1356

```

20 AlAserAlaSerAlaAspProSerLysAspSerLysAlaGlnAlaValSerAl 36
|||||
853 GCTGCGGACCTGCGAGACCGCTCCAGAGACTCCAAAGCTCAGGTTTCTGC 902
|||||
36 AlAlaGlnAlaGlyIleThrGlyThrPTyrAsnGlnLeuGlySerThrP 53
|||||
903 AGCGAAGCTGTATCACTGACCTGTATACCACTGGGGTGCACCTT 952
|||||
53 heIleValThrAlaGlyValaAspGlyAlaLeuThrGlyThrTyrGlnSer 69
|||||
953 TCATTGTGACCGCTGTGTGCGAGAGCTGTGACTGGCGACCTACGAACTT 1002
|||||
70 AlAlaValAlaGlnAlaGlyIleThrGlyThrPTyrAsnGlnLeuGly 86
|||||
1003 GCGGTTGGTAAAGCAATCCGCTACGTAAGTGAAGTGGCGCTTATGACTC 1052
|||||
86 rAlaProAlaThrAspGlyThrAlaLeuGlyThrPTyrValAlaIleT 103
|||||
1053 TGCACCTGCGCACCGATGCTGTGCTACCGCTGTGCGCTGCGCTGCGCTT 1102
|||||
103 rLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSerGlyGln 119
|||||

```

```

1103 GGAACAACTATGCTATGCGCACAGCCGCACTACGTGCTGCCAA 1152
120 TyrValGlyGlyAlaGluAlaArgIleasnThrGlnTrpLeuThrsE 136
1153 TACGTTGGCGGCGCTGAGGCTGCTATCAACACTGAGGCTGTATACATC 1202
136 rGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspT 153
1203 CGGCACACTACGCAAGCATGGAATCGACACTAGTAGTCAATGACA 1252
153 hrpPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLys 169
1253 CATTACCAAAAGTTAAGCTTCTGCTGCTACCATTCGATGCGCAAGAA 1302
170 AlaGlyValaAsnAsnGlyAsnProLeuAspAlaValaGlnGln 183
1303 GCAGGCGTAAACAAAGGTAAACCTCTTACACGCTGTTCAAGCAA 1344

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA2000.DAT:291076
seq_documentation_block:
ID 291076 standard; DNA; 495 BP.
XX
AC 291076;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomyces avidinii prost gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; prost gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomyces avidinii.
XX
PN MO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
DR WPI: 2000-195402/17.
XX
DR P-PSDB; Y80515.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 86-87; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii prost gene as an example of a heterologous gene introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 495 BP; 98 A; 188 C; 151 G; 58 T; 0 other;

```

```

alignment_scores:
    Quality: 838.00      Length: 160
    Ratio: 5.237        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
us-09-589-870-2 x 291076 ..

Align seg 1/1 to: 291076 from: 1 to: 495

24 AlaaspProSerLysAspSerLysAlaGlnAlaSerAlaAlaGlnAlaGln 40
11 GCTACCCCTCCCAAGACTCTGAAGCCCAAGGCTTCGGCGCCGAGGCGCG 60
40 yIleThrGlyThrTrpPyrAsnGlnLeuGlySerThrPheIleValIleThra 57
61 CATCACGCGCACGCTGGTACAAACACAGCTCGGCTCACTTCAATCGTGACCG 110
57 laGlyAlaAspGlyAlaLeuThrGlyThrTrpGlnSerAlaValGlyAsn 73
111 CGGCGCGCGAGCGCGCCCTGACCGGAACCTACGAGTCGCGCGCTGGCAAC 160
74 AlaGlnSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaTh 90
161 GCCGAGAGCCGCTACGCTCGACCGGTCGTTACGACAGCGCGCCGCGCAC 210
90 rAspGlySerGlyThrAlaLeuGlyTTPThrValAlaTrpLysAsnAsnT 107
211 CGAGCGACGCGCGACGCGCCCTCGGTTGACGCGTGGCAAGAAATTAAT 260
107 yTArgAsnAlaHisSerAlaThrTrpSerGlyGlnTyrValGlyGly 123
261 ACCGCAAGCCCACTCCGCAACACGACGCGGACGCGGCACTACGTCGCGCG 310
124 AlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrgl 140
311 GCCGAGGCGAGATCAACACCGATGGCTGCTGACCTCGGACACACCGA 360
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
361 GGCACACGCTGGAAATCAACGCTGGGCGCACGACACCTTCACCAAGG 410
157 aLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValaAsn 173
411 TGAAGCCGTCGCGCCCTCCATCATGACCGCGCAAGAGCGCGCGCTCAAC 460
174 AsnGlyAsnProLeuAspAlaValaGlnGln 183
461 AACGCAACCCGCTCGACGCGCTTCAGACAG 490

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA2000.DAT:291075
seq_documentation_block:
ID 291075 standard; DNA; 492 BP.
XX
AC 291075;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomyces avidinii prost gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; prost gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomyces avidinii.
XX
PN MO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX

```

```

RR 03-AN0-1998: 98IL-0125632.
XX
XX (AGRI-) AGRIC RES ORG.
XX
XX Kapulnk Y, Glnzberg I;
XX
XX WPI: 2000-195402/17.
DR P-PSDB; Y80514.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
XX Examples: Page 86; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidiin prot gene as an example of a heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
XX Sequence 492 BP: 95 A; 188 C; 151 G; 58 T; 0 other:

Alignment_scores:
Quality: 835.00 Length: 161
Ratio: 5.185 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.758

Alignment_block:
US-09-589-870-2 x 291075 ..

Align seg 1/1 to: 291075 from: 1 to: 492

23 SerAlaAspProSerLyAspSerLyAlaGlnValSerAlaAlaGlnVal
:::|||||
5 GCAGTGTGACCCCTCCAAAGACTCGAAGGCCAGGCTCTGGCGCCGAGGC 54

39 aGlyIlethrGlyThrTrpTyraGlnIleuGlySerThrPheIleValT
|||||
55 CGGCTACACGGCGACCTGGTACAAACACACTCGGCTGACCTTCATCTGTA 104

56 hrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrgIuSerAlaValGly 72
|||||
105 CCGCGGGCGCCGACGCGCCCTTGACCGGAACCTACGAGTGGCGGTGGCC 154

73 AsnAlaGluSerArgTyValLeuThrGlyArgTyraSPSerAlaProAl 89
|||||
155 AACGCGAGAGCGCGCTACGTCCTGACCGGCTGTTACGACAGCGCCCGGC 204

89 aThrAspLySerGlyThrAlaLeuGlyTrpThrValAlaIrrPlyAsna 106
|||||
205 CACCCACGCGACGGCGACCGCCCTCGGTGGACCGTGGCGCTGGAAGATA 254

106 snTyraArgAsnAlaHisSerAlaThrThrTrpSPSerGlyGlnTyValGly 122
|||||
255 ACTACCGGACAGCCCGACACTCCGCGACCACTGAGACGGCGCACTACGTGGC 304

123 GlyAlaGlnAlaArgIleAsnThrGlnTrPleuLeuThrSerGlyThrTh 139
|||||
305 GGGCGCGAGCGGAGATCAACACCCAGGCGCTGTCGACCTCGGACACAC 354

139 rGlnAlaAsnAlaIrrPlySerThrIleuValGlyHisAspThrPheThrL 156
|||||
355 CGAGGCGAACGCTGGAAGTCCACGCTGCTGGCGCGACGACACTTCACCA 404

```

```

seq_name: /cgn2_2/gcdata/geneseq/geneseg/NM1988.DAT:N80186
seq_documentation_block:
ID   N80186 standard; DNA; 498 BP.
XX
AC   N80186;
XX
DT   13-OCT-1990 (first entry)
XX
DE   Encodes biosynthetic protein with streptavidin trailer.
XX
KW   Biosynthetic Antibody Binding site (BABS); variable domain;
XX   anti-CEA monoclonal antibody; streptavidin; ss.
XX
OS   Synthetic.
XX
PN   WC8809344-A.
XX
PD   01-DEC-1988.
XX
PF   19-MAY-1988;    88WO-US01737.
XX
PR   21-MAY-1987;    87US-0052800.
XX
PA   (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI   Huston JS, Oppermann H;
XX
DR   WPI: 1988-353928/49.
DR   P-PDB; P80160.
XX
PT   Recombinant multifunctional protein -
XX   having an antibody binding site and a sequence for biological activity
XX   sequestering or binding to a solid support.
PS   Disclosure; ; 115pp; English.
CC   Encodes multi-functional biosynthetic protein comprising single
CC   chain BABS and streptavidin protein trailer linked via a spacer
CC   sequence.
CC   See also NM0171-NM0192.
XX
SO   Sequence 498 BP; 117 A; 137 C; 133 G; 111 T; 0 other;

Alignment_scores:
      Quality: 835.00          Length: 162
      Ratio: 5.154            Gaps: 0
Percent Similarity: 100.000    Percent Identity: 98.148

alignment_block:
US-09-589-870-2 x N80186    ..

Align seg 1/1 to: N80186 from: 1 to: 498

22 AlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaG1 38
.....
4 TC CGGTGGCGACCCGTC CAAGACTCCAAGAAGCTCAAGTTTCTGTCGCGA 53
38 walaGlyIleThrGlyThrTrpTyraAngInLeuNGlySerTrpPheIley 55
|||||
54 AGCTGTAATCATCTGCACCTGGAATACCACTGGGCTGCACATTTCATTG 103
55 alThraAlaGlyAlaAspGlyAlaLeuThrcIyThrGluuSerAlaVal 71

```

```

|||||
104 TGACCGGTGGTGGACGAGACTGTGACTGCGACCTAGCAATCTGGCGTT 153
72 GlysnaIaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaPr 88
154 GGTACCGAGATCCCGCTACGTACTAGCTGGCCGTTATGACTCTGAC 203
88 AlaThrAspGlySerGlyThrAlaLeuGlyTyrThrValAlaTrpLysA 105
204 TGGCACCAGTGGCTGTGTAACCGCTGTGGGCTGGACGTGTGCTTGGAAA 253
105 sNaSnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrVal 121
254 ACAACTATCGTAATGCCACAGCCGCACTACGTGTGCTGGCCAAATAGCTT 303
122 GlyGlyAlaGluAlaArgTyrLeaThrGlnTrpLeuLeuThrSerGlyTh 138
304 GCGGCTGCTGAGCTGTATCAACACACAGTGGCTGTAAACATCGGCGAC 353
138 rThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheT 155
354 TACCGAAGCGATGATGGAATGACACATGATGATGATGATGATGATGATGAT 403
155 hrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGly 171
404 CCAAGTTAAGCTTCTGTGCTGATGATGATGATGATGATGATGATGATGAT 453
172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
454 GTAAACAACGTAACCTCTAGACGCTGTTCAGCAA 489

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA197.DAT:T75491
seq_documentation_block:
ID T75491 standard; DNA; 498 BP.
XX T75491;
XX
XX 27-FEB-1998 (first entry)
DE DNA for streptavidin.
XX
XX S-layer; sbs-A; vaccine; adjuvant; carrier; hybridisation assay;
KW molecular spinning nozzle; molecular laser; streptavidin; ds.
XX
XX Streptomyces avidinii.
OS
XX DE19603649-A1.
XX
XX 07-AUG-1997.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX (LUBI/) LUBITZ W.
XX (SLEY/) SLEYTR U.
XX
XX Kuen B, Lubitz W, Sleytr U;
XX
XX WPI; 1997-394558/37.
XX
XX Preparation of S-layer proteins by expressing sbs-A gene in Gram
XX negative bacterium - or new sbs-B gene in any host, also new
XX recombinant proteins containing heterologous inserts, e.g.
XX epitope(s), useful as vaccines and adjuvants
XX
XX Example 8; Page 26; 31pp; German.
XX
XX S-layer structures comprising the novel Bacillus stearothermophilus
XX PV72 S-layer protein, sbs-A, can be used as vaccines or adjuvants,
XX particularly when they include a bacterial ghost that may contain
XX additional epitopes in its membrane. Other uses of recombinant

```

```

CC sbs-A, depending on the nature of the inserted peptide, are as an
CC universal carrier for biotinylated reactants for use in
CC immunological or hybridisation assays (the insert is streptavidin,
CC preferably encoded by the present sequence), to induce immune
CC responses (epitopes), as a reagent for removing cytokine or toxin
CC from serum (antigenic epitopes), as a molecular spinning nozzle
CC (polyhydroxybutyrate synthase) and as a molecular laser
CC (luciferase).
XX
XX Sequence 498 BP; 120 A; 137 C; 131 G; 110 T; 0 other;
SQ

alignment_scores:
Quality: 834.00 Length: 159
Ratio: 5.245 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x T75491 ..
Align seg 1/1 to: T75491 from: 1 to: 498

25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI 41
7 GACCCGTCGAAGGACTCCAAAGCTCAGGTTCTGCGACCGAAGTGATAT 56
41 eThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValIThrAla 58
57 CACTGGCACCCTGGTATTAACCAACAGTGGGTGACTTTCATTGTAACCCCTG 106
58 lYAlaAspGlyAlaLeuThrGlyTyrGluSerAlaValGlyAsnAla 74
107 GTGCGGACGAGCTCTAGCTAGCGCACTACGAATCTGGCGTTGGTAACGA 156
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
157 GAATCCCGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 206
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyrA 108
207 TGGCTGTGTACCGCTCTGGGCTGGAGCTGTGGCTGTGAAAAAACACTATC 256
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGlyAla 124
257 GTATATGGCAGACGCGCCACTACGTGTGCTGCCAATATGCTGGCGGTGCT 306
125 GluAlaArgTyrLeaThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
307 GAGGCTGTATCAACACTCAGTGGCTGTAAACATCCGCGCACTACCGAGC 356
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158
357 GAATGATGTGAATCGACACTAGTGTGTCATGACACCTTTCACCAAGTTA 406
158 ySPProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAs 174
407 AGCCTTGTGCTAGCATGATGATGCTGCCAAGAAAGCAGCGTTAAACAC 456
175 GlyAsnProLeuAspAlaValGlnGln 183
457 GTTAACCTCTAGACGCTGTTCAGCAA 483

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2001, 21:15:35 ; Search time 62.3 seconds

(without alignments)

199.451 Million cell updates/sec

Title: US-09-589-870-2

Perfect score: 936

Sequence: 1 MRKIVVAIAIVSLTTVSITA.....IDAKKAGVNNGNPLDAVQ 183

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 segs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	936	100.0	183	2	A23513	streptavidin precu
2	932	99.6	183	2	S57284	streptavidin v1 pr
3	897	95.8	183	2	S57285	streptavidin v2 pr
4	146.5	15.7	150	2	S42204	avidin-related pro
5	143.5	15.3	152	1	VICH	avidin precursor [
6	121.5	13.0	150	2	S42201	avidin-related pro
7	120.5	12.9	150	2	S42203	avidin-related pro
8	120	12.8	1064	2	A40136	fibropellin Ia - s
9	116.5	12.4	150	2	S42202	avidin-related pro
10	112	12.0	595	2	A48658	flagellin - Escher
11	101	10.8	595	2	B48658	flagellin - Escher
12	100.5	10.7	1026	2	A48895	paracrystalline su
13	99.5	10.6	1477	2	B43855	high-molecular-wel
14	99.5	10.6	2468	2	A83412	hypothetical prote
15	96.5	10.3	741	2	F70722	probable glcB prot
16	94.5	10.1	507	2	T35677	probable hydrolase
17	94	10.0	570	2	S56132	cellulase (EC 3.2
18	94	10.0	967	2	S66852	hypothetical prote
19	93	9.9	1225	1	B64234	hypothetical prote
20	92.5	9.9	1018	2	H83135	probable adhesin p
21	92	9.8	1217	2	S52714	sericin1B - silkw
22	91.5	9.8	1122	2	G64887	probable tail fibre
23	91.5	9.8	2249	2	A41477	190K surface anti
24	91.5	9.7	867	2	T45463	membrane glycoprot
25	90.5	9.7	2761	2	T21064	hypothetical prote
26	90	9.6	13288	2	T03099	mucin, submaxillar
27	89	9.5	255	1	CPSMUU	muramoylpentapepti
28	89	9.5	488	2	A32235	poly(1,3-hydroxybuty
29	89	9.5	866	2	T45462	membrane glycoprot

30	88.5	9.5	465	2	A47023	S-layer protein -
31	87.5	9.3	457	2	A53669	streptomycies grise
32	87	9.3	666	2	A42296	lysozyme 2 (EC 3.2
33	86	9.2	598	2	T38403	probable nucleopor
34	86	9.2	605	2	T33913	hypothetical prote
35	85.5	9.1	602	2	T45278	oligopeptide ABC t
36	85	9.1	303	2	T35616	probable membrane
37	85	9.1	528	2	S69589	hypothetical prote
38	85	9.1	757	1	VGBEX1	glycoprotein X pre
39	85	9.1	872	2	S49541	cellulase - Cellul
40	85	9.1	1335	2	T30211	autolysin E - Stap
41	85	9.1	1536	2	A43855	high-molecular-wel
42	84.5	9.0	962	2	S03818	carboxymethylcellu
43	83.5	8.9	414	2	A29835	SalI protein - Esc
44	83.5	8.9	490	2	A24262	phase-1 I flagellin
45	83.5	8.9	507	2	T50398	hypothetical serin

ALIGNMENTS

RESULT 1
A23513
streptavidin precursor - Streptomyces avidin1
C:Species: Streptomyces avidin1
C>Date: 03-Nov-1987 #sequence_rev: 103-Nov-1987 #text_change 10-Dec-1999
C:Accession: A23513; S11540
R:Argarana, C.E.; Kuntz, I.D.; Birken, S.; Axel, R.; Cantor, C.R.
Nucleic Acids Res. 14, 1871-1882, 1986
A:Title: Molecular cloning and nucleotide sequence of the streptavidin gene.
A:Reference number: A23513; MID:86148514
A:Accession: A23513
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-183 <ARG>
A:Cross-references: GB:X03591; NID:946740; PIDN:CAA27265.1; PID:946741
R:Gitlin, G.; Bayer, E.A.; Wilchek, M.
Biochem. J. 269, 527-530, 1990
A:Title: Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine res
F:67/Binding site: biotin (Tyr) #status predicted
F:78/Binding site: biotin (Tyr) #status predicted
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>
C:Superfamily: streptavidin
C:Keywords: biotin; homotetramer
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin #status predicted <MAT>
A:Reference number: S11540; MID:90351377
A:Accession: S11540
A:Molecule type: protein
A:Residues: 54-66 'X', 68 <GIT>

S57284
Streptavidin v1 precursor - Streptomyces venezuelae
C:Species: Streptomyces venezuelae
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999
C:Accession: S57284
R:Bayer, E.A.; Kulik, T.; Adar, R.; Walchek, M.
Biochim. Biophys. Acta 1263, 60-66, 1995
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptomyces
A:Reference number: S57284; MUID:95359204
A:Accession: S57284
A:Molecule type: DNA
A:Residues: 1-183 <BAV>
A:Experimental source: Strain Tue2460
C:Superfamily: streptavidin
E:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin v1 #status predicted <MAT>

Query Match	99.6%	Score 932;	DB 2;	Length 183;
Best Local Similarity	99.5%	Pred. No.	1.8e-66;	
Matches 182;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

QY	1	MRIIVAAIAIVSLTITVSTTASASAPSDSKRAQYSAAEAGITGTWNOIGSTPTVYAGAD	60
QY	1	MRIIVAAIAIVSLTITVSTTASASAPSDSKRAQYSAAEAGITGTWNOIGSTPTVYAGAD	60
Db	1	MRIIVAAIAIVSLTITVSTTASASAPSDSKRAQYSAAEAGITGTWNOIGSTPTVYAGAD	60
QY	61	GALTGTYESAVGNAESRYVLTRGRYDAPATPGSGTALGWTGVAKNNVRRNHSATPTMSGOY	12
Db	61	GALTGTYESAVGNAESRYVLTRGRYDAPATPGSGTALGWTGVAKNNVRRNHSATPTMSGOY	12
QY	61	GALTGTYESAVGNAESRYVLTRGRYDAPATPGSGTALGWTGVAKNNVRRNHSATPTMSGOY	12
Db	61	GALTGTYESAVGNAESRYVLTRGRYDAPATPGSGTALGWTGVAKNNVRRNHSATPTMSGOY	12
QY	121	VGGAEARINTOMLLTSGTTEANAMKSTLVGHDPFTTKPKPAASIDAAKAGVNGNPDLA	18
Db	121	VGGAEARINTOMLLTSGTTEANAMKSTLVGHDPFTTKPKPAASIDAAKAGVNGNPDLA	18
QY	121	VGGAEARINTOMLLTSGTTEANAMKSTLVGHDPFTTKPKPAASIDAAKAGVNGNPDLA	18
Db	121	VGGAEARINTOMLLTSGTTEANAMKSTLVGHDPFTTKPKPAASIDAAKAGVNGNPDLA	18

27	101	102	103
	111		
Db	181	V00	183

```

RESULT      3
S57285
streptavidin v2 precursor - Streptomyces venezuelae
C:Species: Streptomyces venezuelae
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999
C:Accession: S57285
R:Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.
Biochim. Biophys. Acta 1263, 60-66, 1995
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom
A:Reference number: S57284; MUID:9535204
A:Accession: S57285
A:Molecule type: DNA
A:Residues: 1-183 <BAV>
A:Experimental source: strain Tue2605
C:Superfamily: streptavidin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin v2 #status predicted <MAT>

```

Query Match	95.8%	Score 897	DB 2	Length 183
Best Local Similarity	95.1%	Pred. No. 1e-63		
Matches 174	Conservative	3	Mismatches	6
			Indels	0
			Gaps	0

[illegible]

QY	181	VQQ	183
Db	181	VQQ	183

RESULT 4
SA2204

avidin-related protein 4/5 precursor - chicken
C;Species: Gallus gallus (chicken)

C:\Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
C:\Accession: S42204; S42205

R. Keinänen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Tolmela, T.A.; Heleni
Eur. J. Biochem. 220, 615-621, 1994

A1: Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1
A1: Reference number: S42201; MUID:94170814

A:Accession: S42204
A:Molecule type: DNA

A;Residues: 1-150 <KEI>
A;Cross-references: EMBL:Z22883; NID:g311811

A;Experimental source: strain White Leghorn; tissue oviduct
A;Genetics: CH1

A;Accession: S42205
A;Molecule type: DN

A;Residues: 1-150 <KEW>
A;Cross-references: EMBL:Z22882; NID:g311812

A;Experimental source: strain White Leghorn; tissue oviduct
A;Genetics: CH2

```
C;Genetics: <CHI>
A;Gene: avr4
```

A; Introns: 27/3; 96/1; 136/2
C; Genetics: <CH2>

A;Gene: AVR5
A;Introns: 27/3; 96/1; 136/2

```

C/superfamily: avian
F:1-24/Domain: signal sequence #status predicted <SIG>

```

F:23-150/Product: avian-related protein 4/5 #status predicted <MATH>
F:28-105/Disulfide bonds: #status predicted

	Query Match	DB 3	Length 100
15 79:	CCCCC TAC E.		

Query match	15.7%	Score 140.5	DB 2	Length 150
Best Local Similarity	33.8%	Pred. 7.5e-05		
Matches 45; Conservative	22;	Mismatches 47;	Indels 19;	Gaps 8;

34 VSAAEAGITGTWYNQLGSTFIYTA-GADGALTGYESAV-----GNAESRYVLTGRYDSAP 88

Db 23 LSARKCSLTGKWTNNLGSIMTIRAVNSRGEFTGYLTAADNPGNITLSPLLGIQHKRA- 811

89 ATDGGTALGWTVAAMKNNYRNAHSATTTWSGQYY---GGAEARINTQWLLTSGTTEAN-AM 144

```

Db      82  ---SQPTGFTVHW--NF--SESTVFTGQCFLDRNGKEV-LKTMMLLRSSVNDISYDW 132

```

QY 145 KSTLVGHDTFTKV 157

```

Db      133 KATRVGYNNFTRL 145
          |||::||::

```

1000

RESULT 5
VICH

avidin precursor [validated] - chicken
C:Species: Gallus gallus (chicken)

C;Date: 24-Apr-1984 #sequence_revision 04-Nov-1994
C;Accession: A54975; A27518; A92093; A92092; A03160

R; Wallén, M.U.; Laakkonen, M.O.; Kulomaa, M.S.
submitted to GenBank, January 1994

A/Description: Sequence of the chicken egg-white avidin gene.
A/Reference number: A54975

A;Accession: A549/5
A;Molecule type: DNA

A;Residues: 1-21,'S',23-152 <WAL>
A;Cross-references: GB:L27818; NID:g450255; PID:g451889

A: Experimental source: adult oviduct, strain White Leghorn
A: Note: difference at position 22 may be due to PCR error in gene sequence

R.Gope, M.L.; Keinonen, R.A.; Kristo, P.A.; Connely, O.M.; Beattie, W.G.; Zarucki-Schu
Nucleic Acids Res. 15, 3595-3606, 1987
A:Title: Molecular cloning of the chicken avidin cDNA.
A:Reference number: A27518; MUID:87203384
A:Accession: A27518
A:Molecule type: mRNA
A:Residues: 1-152 <GDP>
A:Cross-references: GB:X05343; NID:963071; PIDN:CAA28954.1; PID:963072
R.Delange, R.J.; Huang, T.S.
J. Biol. Chem. 246, 686-709, 1971
A:Title: Egg white avidin. III. Sequence of the 78-residue middle cyanogen bromide pep
A:Reference number: A92093; MUID:71107558
A:Accession: A92093
A:Molecule type: protein
A:Residues: 25-57, 'T', 59-76, 'E', 78-152
A:Experimental source: egg white
A:Note: Approximately 50% of the chains have 58-ile
R.Huang, T.S.; Delange, R.J.
J. Biol. Chem. 246, 686-697, 1971
A:Title: Egg white avidin. II. Isolation, composition, and amino acid sequences of the
A:Reference number: A92092; MUID:71107557
A:Accession: A92092
A:Molecule type: protein
A:Residues: 25-57, 'T', 59-76, 'E', 78-152
R.Livnah, O.; Sussman, J.
Submitted to the Brookhaven Protein Data Bank, April 1993
A:Reference number: A51448; PDB:AVT
A:Contents: annotation; X-ray crystallography, 3.0 angstroms, with biotin, residues 27-5
R.Livnah, O.; Bayer, E.A.; Wilchek, M.; Sussman, J.L.
Proc. Natl. Acad. Sci. U.S.A. 90, 5076-5080, 1993
A:Title: Three-dimensional structures of avidin and the avidin-biotin complex.
A:Reference number: A47554; MUID:93281699
A:Contents: annotation; X-ray crystallography, 3.0 angstroms
R.Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M.
Submitted to the Brookhaven Protein Data Bank, March 1993
A:Reference number: A51622; PDB:JAVD
A:Contents: annotation; X-ray crystallography, 2.7 angstroms, with biotin, residues 27-5
R.Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M.
Submitted to the Brookhaven Protein Data Bank, March 1993
A:Reference number: A51623; PDB:JAVE
A:Contents: annotation; X-ray crystallography, 2.8 angstroms, without biotin, residues 2
J. Mol. Biol. 231, 698-710, 1993
A:Title: Three-dimensional structure of the tetragonal crystal form of egg-white avidin
A:Reference number: A54974; MUID:93294833
A:Contents: annotation; X-ray crystallography, 2.7 angstroms
C:Genetics:
A:introns: 27/3; 98/1; 138/2
C:Superfamily: avidin
C:Keywords: glycoprotein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-152/Product: avidin #status experimental <MAT>
F:28-107/Disulfide bonds: #status experimental
F:41/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 15.3%; Score 143.5; DB 1; Length 152;
Best Local Similarity 33.3%; Pred. No. 0.00013;
Matches 44; Conservative 19; Mismatches 56; Indels 13; Gaps 7;

OY 34 VSAAEAGICTGTWNOGSGFTIYVNA--GADGALITYESANGNAESRYVLGRYSAPRTDS 92
DB 23 LLSARKCSLTGKWNDS--IMTIGAVNDGEGFTIYVADNPNGITRSPPLGIQHKR 80
OY 93 SGA--ALGWTVAKKNNYRNAHSAATWSGOV--GGAEARINTQWLLTSGTTE--ANAMKS 146
DB 82 KRQPTFFGFTIVMK-----FSESTIVTTCGCFIDRNGKEV-LKTMILRRSSVNDIGDMKA 136
OY 147 TLVGHDTFTKVK 158
DB 137 TRVGINIFTRLR 148

RESULT 6
S42201
avidin-related protein 1 precursor - chicken
N:Alternate names: avr1
C:Species: Gallus gallus (chicken)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
A:Accession: S42201
R.Keinonen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Heleni
Eur. J. Biochem. 220, 615-621, 1994
A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-
A:Reference number: S42201; MUID:94170814
A:Accession: S42201
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-150 <KEI>
A:Cross-references: EMBL:221611
C:Genetics:
A:introns: 27/3; 96/1; 136/2
C:Superfamily: avidin
C:Keywords: glycoprotein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-150/Product: avidin-related protein 1 #status predicted <MAT>
F:28-105/Disulfide bonds: #status predicted
F:54, 67, 93/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.0%; Score 121.5; DB 2; Length 150;
Best Local Similarity 32.8%; Pred. No. 0.0069;
Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;

OY 34 VSAAEAGICTGTWNOGSGFTIYVNA--DGALITYESAV----GNAESRYVLGTGYDS 86
DB 23 LLSARKCSLTGKWNDS--IMTIGAVNDGEGFTIYVADNPNGITRSPPLGIQHKR 80
OY 87 APATDSCGTALGWTVAKKNNYRNAHSAATWSGO--YV--GGAEARINTQWLLTSGTTE--AN 142
DB 81 A-----CQPTFGFTIVK--NF--SESTIVTTCGCFIDRNGKEV-LKTKWGLRLAVNDISD 130
OY 143 AMKSTLVGHDTFTK 156
DB 131 DMKATRVGNDFTR 144

RESULT 7
S42203
avidin-related protein 3 precursor - chicken
N:Alternate names: avr3 protein
C:Species: Gallus gallus (chicken)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
A:Accession: S42203; S59800
R.Keinonen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Heleni
Eur. J. Biochem. 220, 615-621, 1994
A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-
A:Reference number: S42203; MUID:94170814
A:Accession: S42203
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-150 <KEI>
A:Cross-references: EMBL:221612; NID:965432
R.Kunnas, T.A.; Wallen, M.J.; Kuloma, M.S.
Biochim. Biophys. Acta 1216, 441-445, 1993
A:Title: Induction of chicken avidin and related mRNAs after bacterial infection.
A:Reference number: S39799; MUID:94092737
A:Accession: S39800
A:Status: preliminary; translation not shown
A:Molecule type: mRNA
A:Residues: 71-150 <KUN>
A:Cross-references: EMBL:221536; NID:965429
C:Genetics:
A:Gene: avr3
A:introns: 27/3; 96/1; 136/2
C:Superfamily: avidin

F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-150/Product: avidin-related protein 3 #status predicted <MAT>
F;28-105/Disulfide bonds: #status predicted

Query Match	12.9%	Score 120.5	DB 2	Length 150
Best Local Similarity	30.3%	Pred. No. 0.0082		
Matches	40	Conservative	24	Mismatches 49; Indels 19; Gaps 8;
QY	34	VSAAAGCITGTGTYNOLGSGTIVTA-GADGALGTGESAV----	GNAESRYVLGTRYDSAP	88
	:	: : : : : :		
Db	23	LSARCSILTGKWTNNLGSIMTIRAYNSKGEFGLTLTAVADNPINKISPLGIDHKRR-		81
QY	89	ATDGSITALGWTAVAKNNYRNAASATWSGOYV--GGAEARINTOMLLTSGTE-ANAW		144
	:	: : : :		
Db	82	----CQPFGEGLTWH--NF--SESTSVSEVGGCFIDRSKREY-LTKIKWLQRLAVDIDSIDW		132
QY	145	KSTLVGHDTFK	156	
	:	: :		
Db	133	KATRVGYNNFTR	144	

RESULT 8
A40136
fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)
N:Alternate names: epidermal growth factor homolog precursor
N:Contains: alternatively spliced fibropellin Ib (EGFI)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C>Date: 13-May-1992 #sequence_revision 17-Sep-1997 #text_change 21-Jul-2000
C:Accession: A40136; BA0136; C40136; A29316; A43131
R:Delgallilo-Reynoso, M.G.; Rollo, D.R.; Hersh, D.A.; Raff, R.A.
J. MOL. EVOL. 29, 314-327, 1989
A:Title: Structural analysis of the uEGF gene in the sea urchin Strongylocentrotus purpuratus
#:Reference number: A40136; MUID:50112459

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-114
A:Cross-references: GB:X17530; NID:g10225; PID:9667061
A:Accession: BA0136
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 181-251,329-370,'R',372-408,'RA',411-441 <DE2>
A:Accession: CA0136
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 'K',747-821,898-978 <DE3>
R:Hursh, D.A.; Andrews, M.E.; Raff, R.A.
Science 237, 1487-1490, 1987
A:Title: A sea urchin gene encodes a polypeptide homologous to epidermal growth factor.
A:Reference number: A29316; MUID:87319677
A:Accession: A29316
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'S',280-481,786-1064 <HUR>
A:Cross-references: GB:M1741; NID:g161474; PID:AAA30050.1; PID:g552260
R:Hunt, L.T.; Barker, W.C.
FASEB J. 3, 1760-1764, 1989
A:Title: Avidin-like domain in an epidermal growth factor homolog from a sea urchin.
A:Reference number: A4331; MUID:89196806
A:Contents: annotation
C:Comment: EGF homology repeats 10-17 are spliced out in the short form (fibropellin Ib
C:Superfamily: C1r/C1s repeat homology; EGF homology
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1064/Product: fibropellin I #status predicted <FIR>
F:23-54/Domain: EGF homology <EG01>
F:57-175/Domain: C1r/C1s repeat homology <C1R>
F:180-211/Domain: EGF homology <EG02>
F:218-249/Domain: EGF homology <EG03>
F:256-287/Domain: EGF homology <EG04>
F:294-325/Domain: EGF homology <EG05>
F:332-363/Domain: EGF homology <EG06>
F:370-401/Domain: EGF homology <EG07>

F:408-439/Domain: EGF homology <EG08>
F:446-477/Domain: EGF homology <EG09>
F:484-515/Domain: EGF homology <EG10>
F:522-553/Domain: EGF homology <EG11>
F:560-591/Domain: EGF homology <EG12>
F:598-629/Domain: EGF homology <EG13>
F:636-667/Domain: EGF homology <EG14>
F:674-705/Domain: EGF homology <EG15>
F:712-743/Domain: EGF homology <EG16>
F:750-781/Domain: EGF homology <EG17>
F:788-819/Domain: EGF homology <EG18>
F:826-857/Domain: EGF homology <EG19>
F:864-895/Domain: EGF homology <EG20>
F:902-933/Domain: EGF homology <EG21>
F:936-1064/Region: avdlin-like
F:123-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,261
57,451-466,466-487,484-493,497-542,544-553,560-571,565-580,581-591,598-609,603-618,620
F:489-504,506-515,522-533,527-542,544-553,560-571,565-580,581-591,598-609,603-618,620
08,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/D1

Query	March	12.8%	Score 120;	DB 2;	Length 1064;
Best	Local Similarity	28.4%;	Pred. No. 0.07;		
Matches	40; Conservative	24;	Mismatches	53;	Indels 24; Gaps 7;
QY	34	VSAAEAG--ITGTYWYNDLGSFFIYTAGADGALSTY-----ESAVGNMESRYVL---TGR	83		
		: : : : : : : : : : : : : : : : : : : :			
Db	931	VNCEYVGVCDEEGMYNECNDQVITTKTSTGMLLDYMTYNERALGIAPYVVGVIASNN	990		
QY	84	YDSAPATDGSGLTALGWTYAMKNNYNNAHSATWASGQ--YVGAEARINTQWLITS--GTTEA	141		
		: : : : : : : : : : : : : : : : : : : :			
Db	991	YDF-----PSFGFTYV---RDNGSTTSWTCQCHLDCBEVLTYTTWINTNNVSTIQ	1038		
QY	142	NAKSTLYGHDTFTYKPSAA	162		
		: : : : : : : : : : : : : : : : : : : :			
Db	1039	DIKSSNMVGODKWTTRYEQSIA	1059		

RESULT 9
S42202
avidin-related protein 2 precursor - chicken
N:Alternate names: avr2 protein
C:Species: Gallus gallus (chicken)
C:Date: 13-Jan-1995 #sequence-revision 13-Jan-1995 #text-change 07-May-1999
C:Accession: S42202; S39799
R:Kinnunen, R.A.; Wallén, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Helenius
Eur. J. Biochem. 220. 615-621, 1994
A>Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-
A:Accession: S42202
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-150 <KEI>
A:Cross-references: EMBL:Z21554; NID:g65430
R:Kunнас, T.A.; Wallén, M.J.; Kulomaa, M.S.
Biochim. Biophys. Acta 1216. 441-445, 1993
A>Title: Induction of chicken avidin and related mRNAs after bacterial infection.
A:Reference number: S39799; M01D:94092737
A:Accession: S39799
A:Status: preliminary; translation not shown
A:Molecule type: mRNA
A:Residues: 71-150 <KUN>
A:Cross-references: EMBL:Z21535; NID:g65428
C:Genetics:
A:Gene: avr2
A:Introns: 27/3. 96/1; 136/2
C:Superfamily: avidin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-150/Product: avidin-related protein 2 #status predicted <MAT>
;28-105/Dissulfide bonds: #status predicted

Query Match 12.48; Score 116.5; DB 2; Length 150;

A:Accession: B43855
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1477 <BAR>
 A>Note: sequence inconsistent with the nucleotide translation
 A>Note: sequence extracted from NCBI backbone (NCBIN:89237, NCBI:89240)

Query Match 10.6%; Score 99.5; DB 2; Length 1477;
 Best Local Similarity 26.2%; Pred. No. 4.1;
 Matches 50; Conservative 26; Mismatches 88; Indels 27; Gaps 6;

QY 11 VSLTTSITASASADPSKSKAQSAAEAGI---TGTWYNOL-GSTFTVAGDGLTGT 66
 Db 1177 ISGNVSVSATVLTITKSGSKIEAKSGEANTTSATGTTGCTISGNTVATNA-GDLTVG 1235
 QY 67 YESAVGNMESRYVLTGRDSPAIDGSG-----TALGWTVAWKNNYRNHSAITW 116
 Db 1236 NGAELNATEGAATLTATGNTLTTEAGSSITSTKGVLDLALONGSIAGSINAAVTLMTTG 1295
 QY 117 SCQYVGAEARINTOWLLTSGTTEANAMKSTLVGHDFETTKYKPSA-----ASIDAKKA 170
 Db 1296 TLTIVAGSDIR-----ATSGTLVINAKDAKLNGDASGDSSTEVANVAGSGSVTAATSS 1349
 QY 171 GYNNGNPLDAY 181
 Db 1350 SVNITGDLNTV 1360

RESULT 14

hypoetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
 C:Accession: A83412
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lam, N.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
 A:Reference number: A82950
 A:Accession: A83412
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2468 <STC>
 A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AA05263.1; GSPDB:GN001
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Gene: PA1874

Query Match 10.6%; Score 99.5; DB 2; Length 2468;
 Best Local Similarity 24.9%; Pred. No. 6.9;
 Matches 61; Conservative 19; Mismatches 70; Indels 95; Gaps 12;

QY 14 TTVSITASASA-----DPSKDSKAQVSAEAG-ITGTWYNOLGSTFTVTAAG--- 58
 Db 1381 TVVNVITASDAAGNTSLPATTVDSSLPISIPQVDSNGSVISGT--ADAGNTTIIITDGN 1438
 QY 59 -----ADGA-----LTGTESA VGNMESRYVLTGRDSPA----- 89
 Db 1439 PIGQVTAADSSGNMSTTPGIPLPDGTIVNVNVARSPS-----NVDSPAVITVDGVAPAAP 1492
 QY 90 -----TDGSGTA-LGWTVAWKNNYRNHSAITWVGAAGARINTOWLLTSGTTEA 141
 Db 1493 VIDPSNGTISGTAEGATVIT-----LTDSGAGNPIGQATADGSGNMWTFPTSTPLA 1542
 QY 142 NAKMKTSLVGHDFETTKYK-PSAASIDA-----AKKAGV-----NNGN 176
 Db 1543 NCTVINA VADPAGTSGPASTVVDALAPAPVNPNSNGVVISGTAEGATVITLDGN 1602
 QY 177 PLDAV 181

Db 1603 PIGQV 1607

RESULT 15

F70722
 probable glcB protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: F70722
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, R.; Knäuper, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70722
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-741 <COL>
 A:Cross-references: GB:Z78020; GB:AL123456; NID:g3261625; PIDN:CAB01465.1; PID:e25767
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: glcB

Query Match 10.3%; Score 96.5; DB 2; Length 741;
 Best Local Similarity 26.5%; Pred. No. 3.4;
 Matches 56; Conservative 28; Mismatches 86; Indels 41; Gaps 11;

QY 9 INVSITTSITASASD---PSKDSKAQVSAEAGITGTWYNOLGSTFTVTAAGDGLTGT 65
 Db 100 ITTSGVDAETTTTGAQPLVVPVPLNARFALNANAR-WGSLYDALYGTDIVPE-TDGAERG 157
 QY 66 -TVESA VGNMESRYVLTGRDSPAIDGSG-GTALGWT-----VAKNNYRNHSAITW 116
 Db 158 PTYNKVRGSKVIAVAKKFLDDSVPLSSGSGFDATGTTVDGQLVVALPKSTGLANPGOF 217
 QY 117 SCQYVGAEARINTOWLLTSGTTEANAMKSTLV-GHDFETTKYK 159
 Db 218 AG-YTGAES--PTSVLLNHGILHIEILLIDPESQVGTTRACGVKDYLESATTTIMDFED 274
 QY 160 SAASIDAKKA-----GYNNGNPLDAYQ 183
 Db 275 SYAAVDAADKVLGYRNWGLNKGDLAAAVDK 305

Search completed: April 6, 2001, 00:17:14
 Job time: 10899 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 00:17:16 ; Search time 67.49 Seconds
(without alignments)
87.566 Million cell updates/sec

Title: US-09-589-870-2
Perfect score: 936
Sequence: 1 MRKIVAAVAIVSLTTSITA.....IDAKKAGVNNGNPLDAVQ 183

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwisProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	936	100.0	183	1 SAV_STRV	P22629 streptomyc
2	932	99.6	183	1 SAV1_STRV	Q53533 streptomyc
3	897	95.8	183	1 SAV2_STRV	Q53533 streptomyc
4	146.5	15.7	150	1 AVR4_CHICK	P56734 gallus gall
5	143.5	15.3	152	1 AVTD_CHICK	P02701 gallus gall
6	124.5	13.3	150	1 AVR6_CHICK	P56735 gallus gall
7	121.5	13.0	150	1 AVR1_CHICK	O13153 gallus gall
8	120.5	12.9	150	1 AVR3_CHICK	P56736 gallus gall
9	120.5	12.9	150	1 AVR3_CHICK	P56733 gallus gall
10	120	12.8	1064	1 FBPI_STRPU	P10079 strongyloc
11	116.5	12.4	150	1 AVR2_CHICK	P56732 gallus gall
12	101.5	10.3	1025	1 SLAP_CAUCR	P35828 caulobacter
13	96.5	10.3	741	1 MASZ_MYCTU	Q50596 mycobacteri
14	93	9.9	1225	1 Y309_MYCGE	P47551 mycoplasma
15	91.5	9.8	1120	1 STRF_ECOLI	P76072 escherichia
16	91.5	9.8	2249	1 190K_RICRI	P15921 rickettsia
17	90	9.6	1150	1 APMU_PIG	P12021 sus scrofa
18	89	9.5	255	1 CBPM_STRAL	P00733 streptomyc
19	89	9.5	289	1 HFPD_CAUCR	Q45597 caulobacter
20	89	9.5	488	1 PHB_ALCFR	P12625 allaligenes
21	88.5	9.5	465	1 SLAP_LACBR	Q05044 lacobacilli
22	87.5	9.3	457	1 PRTC_SRRGR	P52320 streptomyc
23	87	9.3	666	1 ALYS_ENTHR	P39046 enterococu
24	86.5	9.2	500	1 ABFL_TTRRE	Q92455 trichoderma
25	86	9.2	598	1 VAUF_SCHPO	Q10168 schizosach
26	85	9.1	797	1 VGLX_HSVB	P28968 equine herp
27	85	9.1	872	1 GUNA_CELFI	P50401 cellulomona
28	84.5	9.0	962	1 GUNA_PSEFL	P10476 pseudomona
29	83.5	8.9	489	1 FLIC_SALTY	P06179 salmonella
30	83.5	8.9	500	1 XYLI_TTRKO	P48792 trichoderma
31	83.5	8.9	575	1 FLA2_CAMJE	P22551 campylobact
32	83.5	8.9	575	1 FLB2_CAMJE	P22552 campylobact
33	83.5	8.9	1645	1 OMPB_RICTY	P96989 r outer mem

34	83	8.9	610	1 CHIT_SRRPL	P11220 streptomyc
35	82.5	8.8	297	1 PRPA_SRRGR	P00776 streptomyc
36	82.5	8.8	331	1 ADPL_WHEAT	Q41629 triticum ae
37	82.5	8.8	563	1 HEMA_IACKV	P09343 influenza a
38	82.5	8.8	616	1 MORA_STRCM	Q05064 streptomyc
39	82	8.8	511	1 GUNB_PSEFL	P18126 pseudomona
40	82	8.8	570	1 FBPI_STRPU	P49013 strongyloc
41	82	8.8	572	1 FLAB_CAMCO	P27053 campylobact
42	82	8.8	572	1 FLAB_CAMCO	P18245 campylobact
43	82	8.8	704	1 OE66_NPVAC	Q00704 autographa
44	82	8.8	1090	1 GUXB_CELFI	P50899 cellulomona
45	82	8.8	2003	1 YDBA_ECOLI	P33666 escherichia

ALIGNMENTS

RESULT 1	ID	SAV_STRV	STANDARD:	PRT:	183 AA.
AC	P22629;	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	STREPTAVIDIN PRECURSOR.				
OS	Streptomyces avidinii.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
RN	(1)				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 24-64.				
RX	MEDLINE-86148514; PubMed-3951999;				
RA	Argarann C.E., Kuntz I.D., Birken S., Axel R., Cantor C.R.;				
RT	"Molecular cloning and nucleotide sequence of the streptavidin gene.";				
RL	Nucleic Acids Res. 14:1871-1882(1986).				
RN	(2)				
RP	IMPORTANCE OF TRP IN BIOTIN-BINDING.				
RX	MEDLINE-89134083; PubMed-3223904;				
RA	Citlin G., Bayer E.A., Wilchek M.;				
RT	"Studies on the biotin-binding site of streptavidin. Tryptophan				
RL	residues involved in the active site.";				
RL	Biochem. J. 256:279-282(1988).				
RN	(3)				
RP	IMPORTANCE OF TYR IN BIOTIN-BINDING.				
RX	MEDLINE-90351377; PubMed-2386489;				
RA	Citlin G., Bayer E.A., Wilchek M.;				
RT	"Studies on the biotin-binding sites of avidin and streptavidin.				
RL	Tyrosine residues are involved in the binding site.";				
RL	Biochem. J. 269:527-530(1990).				
RN	(4)				
RP	CELL-BINDING.				
RX	MEDLINE-90358825; PubMed-2390089;				
RA	Alon R., Bayer E.A., Wilchek M.;				
RT	"Streptavidin contains an RYD sequence which mimics the RGD receptor				
RL	domain of fibronectin.";				
RL	Biochem. Biophys. Res. Commun. 170:1236-1241(1990).				
RN	(5)				
RP	X-RAY CRYSTALLOGRAPHY OF 37-157.				
RX	MEDLINE-89100243; PubMed-2911722;				
RA	Weber P.C., Ohlendorf D.H., Wendoloski J.J., Salem F.R.;				
RT	"Structural origins of high-affinity biotin binding to streptavidin.";				
RL	Science 243:85-88(1989).				
RN	(6)				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.				
RX	MEDLINE-97337436; PubMed-914176;				
RA	Freitag S., Je T'rong I., Klumb L., Stayton P.S., Stenkamp R.E.;				
RT	"Structural studies of the streptavidin binding loop.";				
RL	Protein Sci. 6:1157-1166(1997).				
RN	(7)				
RP	X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS) OF 37-157.				
RX	MEDLINE-97294734; PubMed-9148939;				
RA	Katz B.A., Cass R.T.;				
RT	"In crystals of complexes of streptavidin with peptide ligands				
RT	containing the HPQ sequence the pKa of the peptide histidine is less				

RT than 3.0.";
RL J. Biol. Chem. 272:13220-13228(1997).
RN [8]
RX X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 37-157.
PDB MEDLINE=98070615; PubMed=9405158;
RA Katz B.A.;
RT "binding of biotin to streptavidin stabilizes intersubunit salt
bridges between Asp61 and His87 at low pH.";
RL J. Mol. Biol. 274:776-800(1997).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 37-157 OF MUTANTS.
MEDLINE=98300340; PubMed=9636771;
RA Freitag S., Le Trong I., Chilkoti A., Klumb L.A., Stayton P.S.,
Stenkamp R.E.;
RT "Structural studies of binding site tryptophan mutants in the high-
affinity streptavidin-biotin complex.";
RL J. Mol. Biol. 279:211-221(1998).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -1- DATABASE: NAME=Prozyme technical fact sheet;
WWW="http://www.prozyme.com/technical/sal0data.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
EMBL; X03591; CAA27265.1; -
DR PIR; A23513; A23513.
DR PDB; 1STR; 15-OCT-94.
DR PDB; 1PTS; 15-OCT-94.
DR PDB; 1SRE; 01-NOV-94.
DR PDB; 1SRF; 01-NOV-94.
DR PDB; 1SRG; 01-NOV-94.
DR PDB; 1SRH; 01-NOV-94.
DR PDB; 1SRI; 01-NOV-94.
DR PDB; 1SRJ; 01-NOV-94.
DR PDB; 1RST; 15-MAY-97.
DR PDB; 1RSU; 15-MAY-97.
DR PDB; 1SLD; 03-APR-96.
DR PDB; 1SLE; 03-APR-96.
DR PDB; 1SLF; 03-APR-96.
DR PDB; 1SLG; 03-APR-96.
DR PDB; 1STR; 08-MAR-96.
DR PDB; 1STS; 08-MAR-96.
DR PDB; 1SWA; 04-MAR-98.
DR PDB; 1SWB; 04-MAR-98.
DR PDB; 1SWC; 04-MAR-98.
DR PDB; 1SWD; 04-MAR-98.
DR PDB; 1SWE; 04-MAR-98.
DR PDB; 1SWG; 15-JUL-98.
DR PDB; 1SWH; 23-MAR-99.
DR PDB; 1SWJ; 23-MAR-99.
DR PDB; 1SWK; 23-MAR-99.
DR PDB; 1SWL; 23-MAR-99.
DR PDB; 1SWN; 23-MAR-99.
DR PDB; 1SWO; 23-MAR-99.
DR PDB; 1SWP; 23-MAR-99.
DR PDB; 1SMO; 23-MAR-99.
DR PDB; 1SMR; 23-MAR-99.
DR PDB; 1SWA; 18-MAR-98.
DR PDB; 1SWB; 18-MAR-98.
DR PDB; 1SWC; 18-MAR-98.
DR PDB; 1SWD; 18-MAR-98.
DR PDB; 1SWE; 18-MAR-98.
DR PDB; 1SWF; 18-MAR-98.
DR PDB; 1SWG; 18-MAR-98.

DR PDB; 1WVG; 18-MAR-98.
DR PDB; 1VWH; 18-MAR-98.
DR PDB; 1VWI; 18-MAR-98.
DR PDB; 1VWJ; 18-MAR-98.
DR PDB; 1VWK; 18-MAR-98.
DR PDB; 1VWL; 18-MAR-98.
DR PDB; 1VWM; 18-MAR-98.
DR PDB; 1VWN; 18-MAR-98.
DR PDB; 1VWO; 18-MAR-98.
DR PDB; 1VWP; 18-MAR-98.
DR PDB; 1VWQ; 18-MAR-98.
DR PDB; 1VWR; 18-MAR-98.
DR PDB; 1SWF; 29-APR-98.
DR PDB; 21ZA; 16-SEP-98.
DR PDB; 21ZB; 16-SEP-98.
DR PDB; 21ZC; 16-SEP-98.
DR PDB; 21ZD; 23-SEP-98.
DR PDB; 21ZE; 16-SEP-98.
DR PDB; 21ZF; 16-SEP-98.
DR PDB; 21ZG; 16-SEP-98.
DR PDB; 21ZH; 16-SEP-98.
DR PDB; 21ZI; 16-SEP-98.
DR PDB; 21ZJ; 16-SEP-98.
DR PDB; 21ZK; 16-SEP-98.
DR PDB; 21ZL; 16-SEP-98.
DR PDB; 2RTA; 18-NOV-98.
DR PDB; 2RTB; 18-NOV-98.
DR PDB; 2RTC; 18-NOV-98.
DR PDB; 2RTD; 18-NOV-98.
DR PDB; 2RTE; 18-NOV-98.
DR PDB; 2RTF; 27-JAN-99.
DR PDB; 2RTG; 18-NOV-98.
DR PDB; 2RTH; 18-NOV-98.
DR PDB; 2RTI; 25-NOV-98.
DR PDB; 2RTJ; 25-NOV-98.
DR PDB; 2RTK; 18-NOV-98.
DR PDB; 2RTL; 18-NOV-98.
DR PDB; 2RTM; 18-NOV-98.
DR PDB; 2RTN; 18-NOV-98.
DR PDB; 2RTO; 18-NOV-98.
DR PDB; 2RTP; 18-NOV-98.
DR PDB; 2RTQ; 18-NOV-98.
DR PDB; 2RTR; 18-NOV-98.
DR INTERPRO; IPR000088; -
DR PFAM; PF01382; Avidin; 1.
DR PROSITE; PS00577; Avidin; 1.
KW Signal; Biotin; 3d-structure.
FT SIGNAL 1 24
FT CHAIN 1
FT BINDING 25 183
FT BINDING 67 67
FT BINDING 78 78
FT BINDING 116 116
FT BINDING 132 132
FT BINDING 144 144
FT SITE 83 85
FT SITE 38 41
FT HELIX 3 47
FT STRAND 43 49
FT TURN 48 49
FT STRAND 52 57
FT TURN 59 60
FT STRAND 62 68
FT STRAND 77 84
FT STRAND 95 104
FT STRAND 109 121
FT STRAND 127 136
FT HELIX 140 145
FT STRAND 147 156
SQ SEQUENCE 183 AA; 18834 MW; 4652D8AE018468F5 CRC64;

Query Match 100.0%; Score 936; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 2.7e-69;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
OY 1 MKRIYVAALAVSLTTVSITRNASADPSKDSKAOVSAAEAGITGTWYNQLOSTFTVTAGAD 60
Db 1 MKRIYVAALAVSLTTVSITRNASADPSKDSKAOVSAAEAGITGTWYNQLOSTFTVTAGAD 60
OY 61 GALTGTYESAVNGAESRYVLTGRYDSAPATDGSCTALGWTAVKNNYRNAHSATWWSGOY 120
Db 61 GALTGTYESAVNGAESRYVLTGRYDSAPATDGSCTALGWTAVKNNYRNAHSATWWSGOY 120
OY 121 VGGAERINTQWMLTSGTTEANAMKSTLVGHDTFTYKPSAASIDAKKAGVNNGNPLDA 180
Db 121 VGGAERINTQWMLTSGTTEANAMKSTLVGHDTFTYKPSAASIDAKKAGVNNGNPLDA 180
OY 181 VQO 183
Db 181 VQO 183

RESULT 2
SAV1_STRVL STANDARD: PRT: 183 AA.
AC 053532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE STREPTAVIDIN V1 PRECURSOR (SA V1).
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95359204; PubMed-7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RT from Streptomyces."
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S78777; AAB35015.1; -
CC HSSP: P22629; IPTS.
CC INTERPRO: IPR000088; -.
CC PFAM: PF01382; Avidin; 1.
CC PROSITE: PS00577; AVIDIN; 1.
CC KW Signal; Biotin.
CC FT SIGNAL 1 24 BY SIMILARITY.
CC FT CHAIN 25 183 STREPTAVIDIN V1.
CC FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 78 78 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC SEQUENCE 183 AA; 18864 MW; 4652D8BB14847DE0 CRC64;
FT
SQ
Query Match 99.6%; Score 932; DB 1; Length 183;
```

```
Best Local Similarity 99.5%; Pred. No. 5,7e-69;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 MKRIYVAALAVSLTTVSITRNASADPSKDSKAOVSAAEAGITGTWYNQLOSTFTVTAGAD 60
Db 1 MKRIYVAALAVSLTTVSITRNASADPSKDSKAOVSAAEAGITGTWYNQLOSTFTVTAGAD 60
OY 61 GALTGTYESAVNGAESRYVLTGRYDSAPATDGSCTALGWTAVKNNYRNAHSATWWSGOY 120
Db 61 GALTGTYESAVNGAESRYVLTGRYDSAPATDGSCTALGWTAVKNNYRNAHSATWWSGOY 120
OY 121 VGGAERINTQWMLTSGTTEANAMKSTLVGHDTFTYKPSAASIDAKKAGVNNGNPLDA 180
Db 121 VGGAERINTQWMLTSGTTEANAMKSTLVGHDTFTYKPSAASIDAKKAGVNNGNPLDA 180
OY 181 VQO 183
Db 181 VQO 183
```

```
RESULT 3
SAV2_STRVL STANDARD: PRT: 183 AA.
AC 053533;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE STREPTAVIDIN V2 PRECURSOR (SA V2).
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95359204; PubMed-7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RT from Streptomyces."
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S78782; AAB35016.1; -
CC HSSP: P22629; IPTS.
CC INTERPRO: IPR000088; -.
CC PFAM: PF01382; Avidin; 1.
CC PROSITE: PS00577; AVIDIN; 1.
CC KW Signal; Biotin.
CC FT SIGNAL 1 24 BY SIMILARITY.
CC FT CHAIN 25 183 STREPTAVIDIN V2.
CC FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 78 78 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
CC FT SIMILARITY).
CC SEQUENCE 183 AA; 18833 MW; FEAFFDFE4BDECA CRC64;
FT
SQ
```

Query Match 95.8%; Score 897; DB 1; Length 183;
 Best Local Similarity 95.1%; Pred. No. 3 9e-66;
 Matches 174; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKRIVAIAVSLTFTVITASASADPSKDSKAQVSAEAGITGTWYQLGSTFTVTAGAD 60
 D 1 MKRIVAIAVSLTFTVITASASADPSKDSKAQVSAEAGITGTWYQLGSTFTVTANAD 60
 QY 61 GALTGTESAAGVNAESRYVLTGRDSAPATDGSCTALGTVAKNNRNMAHSATWSGQY 120
 D 61 GSLGTYESAVGNAESRYVLTGRDSAPATDGSCTALGTVAKNNRNMAHSATWSGQY 120
 QY 121 VGGAEARINTOWILTSGTTEANAKSTLVGHDTFTKYKPSAASIDAKKAGVNNGPLDA 180
 D 121 VASGEARINTOWILTSGTTEANAKSTLVGHDTFTKYKPSAASIDAKKAGVNNGPLDA 180
 QY 181 VQQ 183
 D 181 VQQ 183

RESULT 4
 AVID_CHICK STANDARD; PRT; 150 AA.
 AC P56734;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE AVIDIN-RELATED PROTEIN 4/5 PRECURSOR.
 GN AV4 AND AV5.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LEHORN; TISSUE-OVIDUCT;
 RX MEDLINE=94170814; PubMed=8125122;
 RA Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
 RA Toimela T.A., Helenius M.A., Kulomaa M.S.;
 RT "Molecular cloning and nucleotide sequence of chicken avidin-related
 genes 1-5."
 RL Eur. J. Biochem. 220:615-621(1994).
 CC -1- MISCELLANEOUS: THE SEQUENCES OF THE CODING REGIONS OF GENES AV4
 AND AV5 ARE IDENTICAL.
 CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Z22883; -; NOT ANNOTATED_CDS.
 DR INTERPRO: IPR000088; -;
 DR PRINTS: PR00709; AVIDIN.
 DR PROSITE: PS00577; AVIDIN; FALSE_NEG.
 KW Biotin; Signal; Multigene family.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 150 AVIDIN-RELATED PROTEIN 4/5.
 FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 150 AA; 16644 MW; 9AC6C6310EF13A CRC64;

Query Match 15.7%; Score 146.5; DB 1; Length 150;
 Best Local Similarity 33.8%; Pred. No. 1.8e-05;
 Matches 45; Conservative 22; Mismatches 47; Indels 19; Gaps 8;

QY 34 VSAEAGITGTWYQLGSTFTVTA-GADGALTGTYESAV-----GNAESRYVLTGRDSAP 88
 D 23 LSAKCSLITKMTNNGISITFTIRAVNSRGFTGTLYLAVADNPENITLSPLLIQRRA 81
 QY 89 ATDGSCTALGTVAKNNRNMAHSATWSGQY---GGAEARINTOWILTSGTTEAN-AW 144
 D 82 ----SQTFEGFTYHW--NF--SESTVFTGQCFIDRNGKEV-LKTMWLLNSSVNDISYDW 132
 QY 145 KSTLVGHDTFTKY 157
 D 133 KATRVGYNNFTRL 145

RESULT 5
 AVID_CHICK STANDARD; PRT; 152 AA.
 AC P02701; Q91958;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE AVIDIN PRECURSOR.
 GN AVD.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87203384; PubMed=3575102;
 RA Gope M.L., Keinaenen R.A., Kristo P.A., Connely O.M., Beattie W.G.,
 RA Zarnucki-Schulz T., O'Malley B.W., Kulomaa M.S.;
 RT "Molecular cloning of the chicken avidin cDNA."
 RL Nucleic Acids Res. 15:3595-3606(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90355928; PubMed=2143802;
 RA Chandra G., Gray J.G.;
 RT "Cloning and expression of avidin in Escherichia coli."
 RL Meth. Enzymol. 184:70-79(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LEHORN; TISSUE-OVIDUCT;
 RX MEDLINE=95394357; PubMed=7665080;
 RA Wallen M.J., Laukkanen M.O., Kulomaa M.S.;
 RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene
 and its relationship with the avidin-related genes Avr1-Avr5."
 RL Gene 161:205-209(1995).
 RN [4]
 RP SEQUENCE OF 25-152.
 RX MEDLINE=71107558; PubMed=5100763;
 RA Delange R.J., Huang T.-S.;
 RT "Egg white avidin. 3. Sequence of the 78-residue middle cyanogen
 bromide peptide. Complete amino acid sequence of the protein
 subunit."
 RL J. Biol. Chem. 246:698-709(1971).
 RN [5]
 RP IMPORTANCE OF TYR IN BIOTIN-BINDING.
 RX MEDLINE=90351377; PubMed=2386489;
 RA Gitlin G., Bayer E.A., Wilchek M.;
 RT "Studies on the biotin-binding sites of avidin and streptavidin.
 RT Tyrosine residues are involved in the binding site."
 RL Biochem. J. 269:527-530(1990).
 RN [6]
 RP BIOTIN-BINDING STUDIES.
 RX MEDLINE=91378911; PubMed=1898347;
 RA Hiller Y., Bayer E.A., Wilchek M.;
 RT "Studies on the biotin-binding site of avidin. Minimized fragments
 RT that bind biotin."
 RL Biochem. J. 278:573-585(1991).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=93281699; PubMed=8506353;
 RA Livnah O., Bayer E.A., Wilchek M., Sussman J.L.;

```
RT "Three-dimensional structures of avidin and the avidin-biotin  
complex.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:5076-5080(1993).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE=93294833; PubMed=8515446;  
RA Pugliese L., Coda A., Malcovati M., Bolognesi M.;  
RA Silecchi A.G., Pagnelli G., Losso R., Aprida B., Bolognesi M.,  
RA Sidoli A., Arcosi P.;  
RT "Biochemical characterization and crystal structure of a recombinant  
hen avidin and its acidic mutant expressed in Escherichia coli.";  
RL Eur. J. Biochem. 256:453-460(1996).  
CC -1 FUNCTION: THE BIOLOGICAL FUNCTION OF AVIDIN IS NOT KNOWN. FORMS A  
STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE MOLECULE OF  
BIOTIN PER SUBUNIT OF AVIDIN).  
CC -1 TISSUE SPECIFICITY: SYNTHESIZED IN HEN OVIDUCT AND CONCENTRATED IN  
EGG WHITE (WHERE IT REPRESENT 0.05% OF THE TOTAL PROTEIN).  
CC -1 SIMILARITY: BELONGS TO THE AVIDIN/SREPTAVIDIN FAMILY.  
CC -1 DATABASE: NME-Worthington enzyme manual;  
WWW="http://www.worthington-biochem.com/manual/A/Av.html".  
CC CC  
CC -1 DATABASE: NME-Prozyme technical fact sheet;  
WWW="http://www.prozyme.com/technical/avidate.html".  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation at  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC CC  
DR EMBL; X05343; CAAG28954.1; -.  
DR EMBL; L27818; AA859733.1; -.  
DR PIR; A03180; VICH.  
DR PIR; A27518; A27518.  
DR PIR; S11540; S11540.  
DR PDB; ZAVI; 15-JUL-93.  
DR PDB; LAVD; 31-JAN-94.  
DR PDB; IAVE; 31-JUN-94.  
DR PDB; IRAV; 15-JUL-98.  
DR PDB; ZCAM; 15-JUL-98.  
DR INTERPROC: IPROO0088; .  
DR PFAM: PF01382; Avidin; 1.  
DR PRINTS: PR00709; AVIDIN.  
DR PROSITE; PS00577; AVIDIN; 1.  
KW Glycoprotein; Signal; Biotin; 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 25 152  
FT DISULFID 28 107  
FT CARBOHYD 41 41  
FT BINDING 57 57  
FT VARIANT 58 58  
FT CONFLICT 22 22  
FT CONFLICT 77 77  
FT STRAND 32 36  
FT TURN 37 38  
FT STRAND 41 44  
FT TURN 49 50  
FT STRAND 54 58  
FT STRAND 71 74  
FT STRAND 77 77  
FT TURK 83 84  
FT STRAND 87 93  
FT 100 109  
  
N-LINKED (GLCNAC...)  
INVOLVED IN BIOTIN BINDING.  
I-> T (IN APPROX. 50% OF THE CHAINS).  
G-> S (IN REF. 3).  
E-> Q (IN REF. 2 AND 3).
```

FT	STRAND	115	124	
FT	HELIX	130	135	
FT	STRAND	137	146	
SO	SEQUENCE	152 AA;	16769 MW;	1D55A491D5ED5C CRC64;

Query Match	15.3%	Score 143.5;	DB 1;	Length 152;
Best Local Similarity	33.3%	Pred. No. 3.2e-05;		
Matches 44;	Conservative 19;	Mismatches 56;	Indels 13;	Gaps 7;

OY 34 VSAAEAGITGTWYNQIGSTFIYTA-GADGALTGYESAVGNAESRYLVTGRIDSAPATDG 92
 :: : :: | | | : : : | | | : |
Db 23 LSAKCSLTGKWINDGSNMTIGAVNSGEFTGYITAV-TATSNEIKESP LHGTENPIN 81

```

QY 93 SGT--ALGWVAVKNNRYNRAHSATWSQYV--GGAERINTQWLTSCTTE-ANAKKS 146
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 82 KRQPTFEGFTYVNWK----FSESTVFTEOCFIDRNGKEV-LKTMWLILRRSVNDIGDDAKA 136

```

Qy	147	TLVGHDTFTKYK	158
		: :	
Db	137	TRVGINIFTRLR	148

RESULT	6
AVR6_CHICK	
TD	
AVR6_CHICK	
STANDARD:	
PRT:	150 AA.

AC	P56735;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)

DE AVIDIN-RELATED PROTEIN 6 PRECURSOR.
GN AVR6.
OS Gallus gallus (Chicken).
OC Euryarchaeota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Euryarchaeota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;

UC Alcibiosauria; Aves; neoquadraceae; psittaculaceae;
OC Gallus. psittaculaceae;
RN [1] psittaculaceae;
RP SEQUENCE FROM N.A. psittaculaceae;

RT "Characterization and chromosomal localization of the chicken avidin

Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
- I - SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC entifiesrequires a license agreement (See <http://www.isb-slb.ch/announce/>
CC or send an email to license@slb-slb.ch).
CC
CC

DR INTERPRO; IPR000088; -.
DR PRINTS; PR00709; AVIDIN.
DR PROSITE; PS00577; AVIDIN; 1.
DR PDB; 1A6D; Multi-subunit

FT	SIGNAL	1	24	POTENTIAL.
FT	CHAIN	25	150	AVIDIN-RELATED PROTEIN 6.
FT	CARBOHYD	54	50	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	93	93	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	141	141	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	150 AA;	16528 MW;	C0F3357D6291B150 CRC64;

Query Match	13.3%	Score	124.5;	DB	1;	Length	150;
Best Local Similarity	32.8%	Prote.	No. 0.0011;				
Matches	44;	Conservative	25;	Mismatches	42;	Indels	23;
						Gaps	10;

```

34 VSAAEGLIGIMINQLOSGIFIVTGA--DGALIGIIESAV---GMAESKIVLIGRID 88
      :|| : :|| | | ||| | :| | | | :|| :|| :|| :|| :|| :||
09

```

```

Db      23 LSARCSLTGEMDNMLNGS--IMTIGAVNDNGENGNYITAAVNDPENIKLSLPLGIQHR 80
OY      87 AAPATGSGGALGWTYAAKNNRYNAHSATTWQO-YY--GGAERINTQWLTSQTE-AN 142
Db      81 A-----CQPTFFPTVHW--NF--SESTSVFVGQCFVDRSGKEV-LTKTWLQRLAADDISD 130
OY      143 AKWSTLVGHDTFTK 156
Db      131 DMKATRVGYNMFT 144

RESULT 7
AVR1_CHICK
AC      15-JUL-1999 (Rel. 38, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      AVIDIN-RELATED PROTEIN 1 PRECURSOR.
GN      AVR1.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauiria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-OVIDUCT;
RX      MEDLINE=86260103; PubMed=2938690;
RA      kelaenen R.A., Laukkanen M.-L., Kulmaa M.S.;
RT      "Molecular cloning of three structurally related genes for chicken
RT      avidin."
RL      J. steroid Biochem. 30:17-21(1988).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-WHITE LEHORN; TISSUE-OVIDUCT;
RX      MEDLINE=94170814; PubMed=8125122;
RA      Kelaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
RA      Toimela T.A., Helenius M.A., Kulmaa M.S.;
RT      "Molecular cloning and nucleotide sequence of chicken avidin-related
RT      genes 1-5."
RL      Eur. J. Biochem. 220:615-621(1994).
RN      [3]
RP      -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration-
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/isb-sib.ch).
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z21611; -; NOT_ANNOTATED_CDS.
DR      EMBL; Z97063; CAB09798.1; -.
DR      INTERPRO; IPR000088; -.
DR      PRINTS; PR00709; AVIDIN.
DR      PROSITE; PS00577; AVIDIN; 1.
DR      Biotech; Signal; Multigene family.
FT      SIGNAL 1 24
FT      CHAIN 1 24
FT      CARBOYD 25 150
FT      CARBOYD 54 54
FT      CARBOYD 67 67
FT      CARBOYD 93 93
SQ      SEQUENCE 150 AA; 16468 MW; 77548380B995B1P6 CRC64;
-----
Query Match 13.0%; Score 121.5; DB 1; Length 150;
Best Local Similarity 32.8%; Pred. No. 0.0019;
Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;
OY      34 VSAEAGITGMWYNQDSTFTVAG---DGAINGTYSAY---GNASEKRYVLTKGRDS 86
Db      23 LSARCSLTGEMDNMLNGS--IMTIGAVNDNGENGNYITAAVNDPENIKLSLPLGIQHR 80

```

```

QY      87 AARITGSGGALGIMVYAKMKNNRYNAHSATITWSQ-YV--GGAEARINTOMLTISGTE-AN 142
      1  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      81 A-----CQPTFFETVHW--NF--SESTSVFGQCFVDSKGEV-LTKTKLQRLAVIDISD 130
QY      143 AMKSTLVGHDFTFK 156
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      131 DMKATRVGNNDFTK 144

RESULT      8
AVR7_CHICK
ID   AVR7_CHICK          STANDARD:          PRT:   150 AA.
AC   P56736; Q9W6V4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   AVIDIN-RELATED PROTEIN 7 PRECURSOR.
GN   AVR7.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RHOE ISLAND;
RA   Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA   Kuolmaa M.S.;
RT   "Characterization and chromosomal localization of the chicken avidin
RL   gene family."
RL   Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC   -I- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
-----
CC
DR   EMBL; AJ237659; CAB93894.1; -.
DR   INTERPRO; IPR000088; -.
DR   PRINTS; PR00709; AVIDIN.
DR   PROSITE; PS00577; AVIDIN; 1.
KM   Biotin; Signal; Multigene family.
FT   SIGNAL              1
FT   CHAIN              124
FT   POTENTIAL.
FT   AVIDIN-RELATED PROTEIN 7.
FT   CARBOHYD            41      41      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD            54      54      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD            93      93      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD            141     141     N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE            150 AA; 16543 MW; 30F35611F65923D CRC64;

```

	Query Match	13.0%	Score 121.5;	DB 1;	Length 150;	
	Best Local Similarity	31.1%;	Pred. No. 0.0019;			
Matches	41;	Conservative	23;	Mismatches	45;	Indels 19; Gaps 8;
QY	34 VSAAEAGITGTWYNOLSGSTFIYTAGAD-GALTGYESAV----GNAESRYVLTRYDSAP	88				
	: :: : :		:	:	:	:::
Dd	23 LSARCSLTLEMDNNLGSNMTIGAVNDNGEEFGTYITTVADNPGRNIKSLPLGIQHKA-	81				
QY	89 ATDGGCTALGWIVAMKKNVRRNAHSHTWSGQY---GCAGEAINTOWLLTSCTTE-ANAW	144				
	: :	:	:	:	:	:
Dd	82 ----CQTFEGFLFVHH--NF--SESTSVEFGCCFDIRSGREY-LKTWKWLORLAVIDSDIW	132				
QY	145 KSTLVGHDTFTK	156				
	: : :: :					
Dd	133 KATRVGVNNPFR	144				
RESULT	9					
AVR3_CHICK						

```

ID AVB3_CHICK STANDARD; PRT; 150 AA.
AC P56733;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AVIDIN-RELATED PROTEIN 3 PRECURSOR.
GN AVR3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-OVIDUCT;
RX MEDLINE-94170814; PubMed-8125122;
RA Keiminen R.A., Wallen M.J., Kristo P.A., Laakkonen M.O.,
RA Toimela T.A., Helenius M.A., Kulomaa M.S.;
RT "Molecular cloning and nucleotide sequence of chicken avidin-related
RT genes 1-5."
RL Eur. J. Biochem. 220:615-621(1994).
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 221612; -; NOT_ANNOTATED_CDS.
DR EMBL: 221536; -; NOT_ANNOTATED_CDS.
DR INTERPRO: IPR000088; -.
DR PRINTS: PR00709; AVIDIN.
DR PROSITE: PS00577; AVIDIN.
DR KMW Biotin; signal; multigene family.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 150 AVIDIN-RELATED PROTEIN 3.
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 150 AA; 16657 MW; 84DC03926A6BE21C CRC64;

Query Match 12.9%; Score 120.5; DB 1; Length 150;
Best Local Similarity 30.3%; Pred. No. 0.0023;
Matches 40; Conservative 24; Mismatches 49; Indels 19; Gaps 8;

OY 34 VSAEAGITGTWYNQLGSTFTVTA-GADGALITGYESAV-----GNAESRYVLTGRYSAP 88
DB 23 LSAKRCGLTKMNNNGSMTIRAVNRSGEFAGTYLTAVDNCGNIKLSPLGLGQHKRA- 81
OY 89 ATDGSGLALGWTVAWKNKNYNAHSATWSGOYV---GGAERINTOMLLTSGTTE-ANAW 144
DB 82 ----CQPTFGFTVHM--NF--SESTSVFVGQCFIDRSKGEV-LKTKWLGRLAVDIDSDW 132
OY 145 KSTLVGHDTFTK 156
DB 133 KATRVGYNNFTR 144

RESULT 10
FBP1_STRPU STANDARD; PRT; 1064 AA.
ID FBP1_STRPU
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FIBROPELIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE (UEGF-1).
GN EGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae;

```

```

OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90112459; PubMed-2514273;
RA Delgado-Illo-Reynoso M.G., Rolio D.R., Hursh D.A., Rafi R.A.;
RT "Structural analysis of the UEGF gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats."
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE-87319677; PubMed-3498216;
RA Hursh D.A., Andrews M.E., Rafi R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor."
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE-89196806; PubMed-2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin."
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE-91285254; PubMed-2060714;
RA Bisgrove B.W., Andrews M.E., Rafi R.A.;
RT "Fibroblasts, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo."
RL Dev. Biol. 146:89-99(1991).
CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
CC REPEATS.
CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08692; AAA62164.1; -
DR EMBL: L08692; AAA62163.1; -
DR EMBL: X17530; CAA35571.1; -
DR EMBL: M17421; AAA30050.1; -
DR EMBL: X17533; CAA35573.1; -
DR PIR: A29316; A29316.
DR HSSP: P01132; IEPH.
DR INTERPRO: IPR000088; -.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR000859; -.
DR INTERPRO: IPR001438; -.
DR INTERPRO: IPR001881; -.
DR PFAM: PF01382; Avidin; 1.
DR PFAM: PF00431; CUB; 1.

```

FT	DISULFID	620	629	BY SIMILARITY.
FT	DISULFID	636	647	BY SIMILARITY.
FT	DISULFID	641	656	BY SIMILARITY.
FT	DISULFID	658	667	BY SIMILARITY.
FT	DISULFID	674	685	BY SIMILARITY.
FT	DISULFID	679	694	BY SIMILARITY.
FT	DISULFID	696	705	BY SIMILARITY.
FT	DISULFID	712	723	BY SIMILARITY.
FT	DISULFID	717	732	BY SIMILARITY.
FT	DISULFID	734	743	BY SIMILARITY.
FT	DISULFID	750	761	BY SIMILARITY.
FT	DISULFID	755	770	BY SIMILARITY.
FT	DISULFID	772	781	BY SIMILARITY.
FT	DISULFID	788	799	BY SIMILARITY.
FT	DISULFID	793	808	BY SIMILARITY.
FT	DISULFID	810	819	BY SIMILARITY.
FT	DISULFID	826	837	BY SIMILARITY.
FT	DISULFID	902	913	BY SIMILARITY.
FT	DISULFID	907	922	BY SIMILARITY.
FT	DISULFID	924	933	BY SIMILARITY.
FT	CARBOHYD	30	30	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	136	136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VASAPLIC	477	780	MISSING (IN ISOFORM 1B).
FT	CONFLICT	279	279	L -> S (IN REF. 2).
SO	SEQUENCE	1064 AA;	112072 MM;	25569CA012ED6D09 CRC64;

Query Match	12.8%;	Score 120;	DB 1;	Length 1064;
Best Local Similarity	28.4%;	Pred. No. 0.02;	Mismatches	24;
Matches	40;	Conservative	24;	Gaps

Qy	34	VSAAEAG---	ITTGWYNOLGSTFTAGAGALGTGY---	ESAVGNABSRVLT--	TGR	83
Db	931	VNCEVEGFCDEGMMWYEDCNDQYITITSGMMIGDMYTNENALTYAAPTYYVYGASNN				990
Qy	84	YDSAPATDGSGLATGWTVAAMKNTRNNAASFTWSGO--	YVGGAEARINTQMLTTS--	GTTEA		141
Db	991	YDF-----	PSRGFTVV----	RDNQGSFTWTCQCHLDGEGVELYTTWIMTNMSTQ		1039
Qy	142	NAMKSTLVGHDTFKVPSAA				162
Db	1039	DIKSNMVGODKMTREOSTA				1059

RESULT	11
AVR2_CHICK	
ID	AVR2_CHICK
AC	P56732;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	AVIDIN-RELATED PROTEIN 2 PRECURSOR.
GN	AVR2.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-WHITE LEGHORN; TISSUE=OVIDUCT;
RX	MEDLINE=94170814; PubMed=8125122;
RA	Kainanen R.A., Kallen M.J., Kristo P.A., Laukkanen M.O.,
RA	Toimela T.A., Helenius M.A., Kulomaa M.S.;
RT	Molecular cloning and nucleotide sequence of chicken avidin-related
RT	genes 1-5."*
RL	Eur. J. Biochem. 220:615-621(1994).
CC	-1- SIMILARITY. BELONGS TO THE AVIDIN/SPEP/AVIDIN FAMILY


```

RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RL      Nature 393:537-544(1998).
CC      -1- CATABOLIC ACTIVITY: L-MALATE + COA = ACETYL-COA + H(2)O +
CC      GLYOXALATE.
CC      -1- PATHWAY: SECOND STEP IN GLYOXALATE BYPASS, AN ALTERNATIVE TO THE
CC      TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).
CC      -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE MALATE SYNTHASE G FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: Z78020; CAB01465.1; -.
KW      TUBERCULIST; RV1837C; -.
KW      Glyoxylate bypass; Tricarboxylic acid cycle; Lyase.
SQ      SEQUENCE 741 AA; 80403 MW; A92F54E0FE8B7C64 CRC64;

Query Match
Best Local Similarity 10.3%; Score 96.5; DB 1; Length 741;
Matches 56; Conservative 28; Mismatches 86; Indels 41; Gaps 11;

OY      9 IAVSLTTSITASASAD---PSKSKAQSAAEAGITGTWYNOLGSTFTVTAGADALNG 65
DB      100 ITTSGVDAETTTAGPQVLPVLANRFAALNANAR-WGSLYDLVGTDTVPE-TDGAENG 157
OY      66 -TYESAVGNAESRYVTGKYDAPATDGS-CTALGWT-----VAKNNRYNAHSATTW 116
DB      158 PYYNKVRGKVIAYARKLDDSVLSSGSGFATGFTVDGOLVVALPDKSGLANPGOF 217
OY      117 SGQYVGAGARINTOWLTS-----GTTEANMKSTLV--GHDTFKKVP 159
DB      218 AG-YTGAAS--PTSYLLNHGHLHITLIDPESQVGTIDRAGKVDILSATTTIMDFED 274
OY      160 SASASIDAKKA-----GVNNGNPLDAVQO 183
DB      275 SVAAYDAADKVLGYRNLGLINKGDLAAAVDK 305

RESULT 14
Y309_MYCGE STANDARD; PRT; 1225 AA.
ID      Y309_MYCGE
AC      P47551; 049317;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOTHETICAL LIPOPROTEIN MG309 PRECURSOR.
GN      MG309.
OS      Mycoplasma genitalium.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC      Mycoplasmataceae; Mycoplasma.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 33530 / G-37;
RX      MEDLINE=96026346; PubMed=756993;
RA      Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA      Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA      Fritchman J.L., Weidman J.F., Small K.V., Phillips M., Fuhmann J.L.,
RA      Nguyen D.T., Usterback T.R., Saudes D.M., Sandilands C.A., Merrick J.M.,
RA      Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lueker T.S.,
RA      Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT      "The minimal gene complement of Mycoplasma genitalium.";
RL      Science 270:397-403(1995).
RN      [2]
RP      SEQUENCE OF 1138-1224 FROM N.A.
RC      STRAIN-ATCC 33530 / G-37;

```

```

RX      MEDLINE=94075230; PubMed=8253680;
RA      Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III.
RT      "A survey of the Mycoplasma genitalium genome by using random
RT      sequencing."
RL      J. Bacteriol. 175:7918-7930(1993).
CC      -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC      ANCHOR (POTENTIAL).
CC      -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U39711; AAC71531.1; -.
DR      EMBL: U02200; AAD12488.1; -.
DR      TIGR: MG309; -.
DR      PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Hypothetical protein; Lipoprotein; Membrane; signal.
FT      SIGNAL 1 27
FT      CHAIN 28 1225
FT      LIPID 28 28
FT      CONFLICT 1185 1185 L->V (IN REF.2).
SQ      SEQUENCE 1225 AA; 138375 MW; C3E4BF5B4319B6E8 CRC64;

Query Match
Best Local Similarity 9.9%; Score 93; DB 1; Length 1225;
Matches 56; Conservative 22; Mismatches 78; Indels 72; Gaps 9;

OY      4 IYVAALVSLTTSITASASADPS-----KSKAQSAAEAGITGTWY---NOLGS 51
DB      12 LLTSSIAVSL--GIYAVACAPNSRTIENLFRPSSAFDKNDGSIATVLYALENREGL 68
OY      52 TFLVTAGAGALGTYESAVGNAESRYVTGKYDAPATDGSCTALGWTYVAKNNRYNANH 111
DB      69 TQYLWRLAPVLRNFEENVDDIKRNLFTFTMD--TDNS-----FVNOEONLRN-- 116
OY      112 SATWMSGQYVGAGARINTOWLTS-----HDTEFKKVP 156
DB      117 -----QYRGDIYVRLQDIIIDNTGNGAN-WKLRDVANNKIYDDEINKLFRNVEYVD 168
OY      157 -----VKPSASIDAKKAGVNNGNPLDAVQ 182
DB      169 KSYGVLTPLKGLIENOSNMNMIKIOAKFVDKRRKLRLINNDVYAAIQ 216

RESULT 15
STFR_ECOLI STANDARD; PRT; 1120 AA.
ID      STFR_ECOLI
AC      P76072; P77560;
DT      01-OCT-2000 (Rel. 40, Created)
DT      01-OCT-2000 (Rel. 40, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC.
GN      STFR OR B1372.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [2]

```


RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.,
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000234; AAC74454.1; ALT_INIT.
DR EMBL: D90774; BAA14966.1; -
DR EMBL: D90775; BAA14975.1; -
DR ECOGENE: EG13370; STFR.
DR HSSP: P04002; 1MFA.
DR INTERPRO: IPR000122; -
DR Hypothetical protein; Fiber protein; Repeat.
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

Query Match 9.8%; Score 91.5; DB 1; Length 1120;
Best Local Similarity 24.7%; Pred. No. 4.2; Mismatches 89; Indels 7; Gaps 4;
Matches 42; Conservative 32;
OY 10 AVSLTVSITASADPPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGAGALTGYES 69
DB 219 ATASATITATASDAATSARDAASKEAKSSETNA--SSASSASASATAGNSAKRAKT 276
OY 70 AVGNASRYVLTGRYDSAPATDGGTALGWTV-AMKNNYRNAHSATFTWSGQYVGAEART 128
DB 277 SETNARSSETTAAGQ--SASAAGSKTAAASSASAASSTAGQASASATAGKSAESAASSA 334
OY 129 NTQWLTSQTTE--ANAKMSTLVGHDTFTYVKPSAASIDAAKAGVANGN 176
DB 335 STATTRAGATEQASAAARSASAKTSETNAKASETSAESSTAAASSAS 384

Search completed: April 6, 2001, 00:19:51
Job time: 155 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 00:16:01 ; Search time 66.64 Seconds
(without alignments)
321.864 Million cell updates/sec

Title: US-09-589-870-2
Perfect score: 936
Sequence: 1 MRKIVVAIAVSLFTVSITA.....IDAAKKAGVNGNPLDAVQQ 183

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	114	12.2	529	5	025058 heliocidari
2	112	12.0	595	2	006335 escherichia
3	110.5	11.8	892	2	09kx38 moraxella c
4	105.5	11.3	1001	3	005164 saccharomyc
5	104	11.1	595	2	095077 escherichia
6	101	10.8	595	2	006351 escherichia
7	100.5	10.7	1026	2	09RF12 caulobacter
8	100	10.7	565	2	09S4M2 escherichia
9	100	10.7	565	2	09S4M4 escherichia
10	99.5	10.6	1477	2	048028 haemophilus
11	98.5	10.5	832	2	054356 moraxella c
12	98	10.5	595	2	09S0U1 escherichia
13	97.5	10.4	729	2	09KH34 antarctic b
14	97	10.4	443	6	028243 canis fami1
15	94.5	10.1	507	2	050506 streptomyce
16	94.5	10.1	381	2	09KX33 streptococ
17	94	10.0	370	2	059665 pseudomonas
18	94	10.0	967	3	008294 saccharomyc
19	92	9.8	865	2	043919 aeromonas c

20	92	9.8	1217	5	017240 bombyx mori
21	91.5	9.8	922	2	09L964 moraxella c
22	91.5	9.8	1058	2	P96141 thal tick t
23	91.5	9.8	1058	2	09Z1U3 rickettsia
24	91.5	9.8	1060	2	P95642 rickettsia
25	91	9.7	231	2	09KXK3 streptococ
26	91	9.7	700	2	030678 xanthomonas
27	91	9.7	867	12	039782 equine herp
28	90.5	9.7	568	2	053789 shigella bo
29	90.5	9.7	568	2	09S0T5 escherichia
30	90.5	9.7	1060	2	09LARI rickettsia
31	90.5	9.7	1060	2	09LARO rickettsia
32	90.5	9.7	2178	2	09KWR3 streptococ
33	90.5	9.7	2761	5	019522 caenorhabd1
34	90	9.6	873	2	09L960 moraxella c
35	90	9.6	1183	2	09ROB7 cellulomona
36	90	9.6	13288	6	018758 sus scrofa
37	89	9.5	289	2	045977 caulobacter
38	89	9.5	866	12	039781 equine herp
39	88.5	9.5	1061	2	P95586 rickettsia
40	88.5	9.5	2021	2	052657 rickettsia
41	88	9.4	894	2	09L962 moraxella c
42	88	9.4	1060	2	09LAK7 rickettsia
43	87.5	9.3	470	5	09U6M5 glibodera t
44	87.5	9.3	574	2	09RPY6 campylobact
45	87.5	9.3	1060	2	P95620 rickettsia

ALIGNMENTS

RESULT 1
ID 025058 PRELIMINARY; PRT; 529 AA.
AC 025058;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE FIROPELIN IA (FRAGMENT).
OS Helicoidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinozoa; Echinodermata; Echinodermata; Echinodermata;
OC Helicoidaris.
OX NCBI_TaxID=7634;
RN
RP
RA
RA BLISGROVE B.W.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L33861; AAA2995.1; -.
DR HSSP; P08709; 1BF9.
DR INTERPRO; IPR000083; -.
DR INTERPRO; IPR000088; -.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001010; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR001947; -.
DR INTERPRO; IPR002049; -.
DR PEAM; PF00008; EGF_10.
DR PFAM; PF01382; Avidin; 1.
DR PRINTS; PRO0010; EGF10.
DR PRINTS; PRO0011; EGF10.
DR PRINTS; PRO0012; EGF10.
DR PRINTS; PRO0013; EGF10.
DR PRINTS; PRO0014; EGF10.
DR PRINTS; PRO0015; EGF10.
DR PRINTS; PRO0016; EGF10.
DR PRINTS; PRO0017; EGF10.
DR PRINTS; PRO0018; EGF10.
DR PRINTS; PRO0019; EGF10.
DR PRINTS; PRO0020; EGF10.
DR PRINTS; PRO0021; EGF10.
DR PRINTS; PRO0022; EGF10.
DR PRINTS; PRO0023; EGF10.
DR PRINTS; PRO0024; EGF10.
DR PRINTS; PRO0025; EGF10.
DR PRINTS; PRO0026; EGF10.
DR PRINTS; PRO0027; EGF10.
DR PRINTS; PRO0028; EGF10.
DR PRINTS; PRO0029; EGF10.
DR PRINTS; PRO0030; EGF10.
DR PRINTS; PRO0031; EGF10.
DR PRINTS; PRO0032; EGF10.
DR PRINTS; PRO0033; EGF10.
DR PRINTS; PRO0034; EGF10.
DR PRINTS; PRO0035; EGF10.
DR PRINTS; PRO0036; EGF10.
DR PRINTS; PRO0037; EGF10.
DR PRINTS; PRO0038; EGF10.
DR PRINTS; PRO0039; EGF10.
DR PRINTS; PRO0040; EGF10.
DR PRINTS; PRO0041; EGF10.
DR PRINTS; PRO0042; EGF10.
DR PRINTS; PRO0043; EGF10.
DR PRINTS; PRO0044; EGF10.
DR PRINTS; PRO0045; EGF10.
DR PRINTS; PRO0046; EGF10.
DR PRINTS; PRO0047; EGF10.
DR PRINTS; PRO0048; EGF10.
DR PRINTS; PRO0049; EGF10.
DR PRINTS; PRO0050; EGF10.
DR PRINTS; PRO0051; EGF10.
DR PRINTS; PRO0052; EGF10.
DR PRINTS; PRO0053; EGF10.
DR PRINTS; PRO0054; EGF10.
DR PRINTS; PRO0055; EGF10.
DR PRINTS; PRO0056; EGF10.
DR PRINTS; PRO0057; EGF10.
DR PRINTS; PRO0058; EGF10.
DR PRINTS; PRO0059; EGF10.
DR PRINTS; PRO0060; EGF10.
DR PRINTS; PRO0061; EGF10.
DR PRINTS; PRO0062; EGF10.
DR PRINTS; PRO0063; EGF10.
DR PRINTS; PRO0064; EGF10.
DR PRINTS; PRO0065; EGF10.
DR PRINTS; PRO0066; EGF10.
DR PRINTS; PRO0067; EGF10.
DR PRINTS; PRO0068; EGF10.
DR PRINTS; PRO0069; EGF10.
DR PRINTS; PRO0070; EGF10.
DR PRINTS; PRO0071; EGF10.
DR PRINTS; PRO0072; EGF10.
DR PRINTS; PRO0073; EGF10.
DR PRINTS; PRO0074; EGF10.
DR PRINTS; PRO0075; EGF10.
DR PRINTS; PRO0076; EGF10.
DR PRINTS; PRO0077; EGF10.
DR PRINTS; PRO0078; EGF10.
DR PRINTS; PRO0079; EGF10.
DR PRINTS; PRO0080; EGF10.
DR PRINTS; PRO0081; EGF10.
DR PRINTS; PRO0082; EGF10.
DR PRINTS; PRO0083; EGF10.
DR PRINTS; PRO0084; EGF10.
DR PRINTS; PRO0085; EGF10.
DR PRINTS; PRO0086; EGF10.
DR PRINTS; PRO0087; EGF10.
DR PRINTS; PRO0088; EGF10.
DR PRINTS; PRO0089; EGF10.
DR PRINTS; PRO0090; EGF10.
DR PRINTS; PRO0091; EGF10.
DR PRINTS; PRO0092; EGF10.
DR PRINTS; PRO0093; EGF10.
DR PRINTS; PRO0094; EGF10.
DR PRINTS; PRO0095; EGF10.
DR PRINTS; PRO0096; EGF10.
DR PRINTS; PRO0097; EGF10.
DR PRINTS; PRO0098; EGF10.
DR PRINTS; PRO0099; EGF10.
DR PRINTS; PRO0100; EGF10.
DR PRINTS; PRO0101; EGF10.
DR PRINTS; PRO0102; EGF10.
DR PRINTS; PRO0103; EGF10.
DR PRINTS; PRO0104; EGF10.
DR PRINTS; PRO0105; EGF10.
DR PRINTS; PRO0106; EGF10.
DR PRINTS; PRO0107; EGF10.
DR PRINTS; PRO0108; EGF10.
DR PRINTS; PRO0109; EGF10.
DR PRINTS; PRO0110; EGF10.
DR PRINTS; PRO0111; EGF10.
DR PRINTS; PRO0112; EGF10.
DR PRINTS; PRO0113; EGF10.
DR PRINTS; PRO0114; EGF10.
DR PRINTS; PRO0115; EGF10.
DR PRINTS; PRO0116; EGF10.
DR PRINTS; PRO0117; EGF10.
DR PRINTS; PRO0118; EGF10.
DR PRINTS; PRO0119; EGF10.
DR PRINTS; PRO0120; EGF10.
DR PRINTS; PRO0121; EGF10.
DR PRINTS; PRO0122; EGF10.
DR PRINTS; PRO0123; EGF10.
DR PRINTS; PRO0124; EGF10.
DR PRINTS; PRO0125; EGF10.
DR PRINTS; PRO0126; EGF10.
DR PRINTS; PRO0127; EGF10.
DR PRINTS; PRO0128; EGF10.
DR PRINTS; PRO0129; EGF10.
DR PRINTS; PRO0130; EGF10.
DR PRINTS; PRO0131; EGF10.
DR PRINTS; PRO0132; EGF10.
DR PRINTS; PRO0133; EGF10.
DR PRINTS; PRO0134; EGF10.
DR PRINTS; PRO0135; EGF10.
DR PRINTS; PRO0136; EGF10.
DR PRINTS; PRO0137; EGF10.
DR PRINTS; PRO0138; EGF10.
DR PRINTS; PRO0139; EGF10.
DR PRINTS; PRO0140; EGF10.
DR PRINTS; PRO0141; EGF10.
DR PRINTS; PRO0142; EGF10.
DR PRINTS; PRO0143; EGF10.
DR PRINTS; PRO0144; EGF10.
DR PRINTS; PRO0145; EGF10.
DR PRINTS; PRO0146; EGF10.
DR PRINTS; PRO0147; EGF10.
DR PRINTS; PRO0148; EGF10.
DR PRINTS; PRO0149; EGF10.
DR PRINTS; PRO0150; EGF10.
DR PRINTS; PRO0151; EGF10.
DR PRINTS; PRO0152; EGF10.
DR PRINTS; PRO0153; EGF10.
DR PRINTS; PRO0154; EGF10.
DR PRINTS; PRO0155; EGF10.
DR PRINTS; PRO0156; EGF10.
DR PRINTS; PRO0157; EGF10.
DR PRINTS; PRO0158; EGF10.
DR PRINTS; PRO0159; EGF10.
DR PRINTS; PRO0160; EGF10.
DR PRINTS; PRO0161; EGF10.
DR PRINTS; PRO0162; EGF10.
DR PRINTS; PRO0163; EGF10.
DR PRINTS; PRO0164; EGF10.
DR PRINTS; PRO0165; EGF10.
DR PRINTS; PRO0166; EGF10.
DR PRINTS; PRO0167; EGF10.
DR PRINTS; PRO0168; EGF10.
DR PRINTS; PRO0169; EGF10.
DR PRINTS; PRO0170; EGF10.
DR PRINTS; PRO0171; EGF10.
DR PRINTS; PRO0172; EGF10.
DR PRINTS; PRO0173; EGF10.
DR PRINTS; PRO0174; EGF10.
DR PRINTS; PRO0175; EGF10.
DR PRINTS; PRO0176; EGF10.
DR PRINTS; PRO0177; EGF10.
DR PRINTS; PRO0178; EGF10.
DR PRINTS; PRO0179; EGF10.
DR PRINTS; PRO0180; EGF10.
DR PRINTS; PRO0181; EGF10.
DR PRINTS; PRO0182; EGF10.
DR PRINTS; PRO0183; EGF10.
DR PRINTS; PRO0184; EGF10.
DR PRINTS; PRO0185; EGF10.
DR PRINTS; PRO0186; EGF10.
DR PRINTS; PRO0187; EGF10.
DR PRINTS; PRO0188; EGF10.
DR PRINTS; PRO0189; EGF10.
DR PRINTS; PRO0190; EGF10.
DR PRINTS; PRO0191; EGF10.
DR PRINTS; PRO0192; EGF10.
DR PRINTS; PRO0193; EGF10.
DR PRINTS; PRO0194; EGF10.
DR PRINTS; PRO0195; EGF10.
DR PRINTS; PRO0196; EGF10.
DR PRINTS; PRO0197; EGF10.
DR PRINTS; PRO0198; EGF10.
DR PRINTS; PRO0199; EGF10.
DR PRINTS; PRO0200; EGF10.
DR PRINTS; PRO0201; EGF10.
DR PRINTS; PRO0202; EGF10.
DR PRINTS; PRO0203; EGF10.
DR PRINTS; PRO0204; EGF10.
DR PRINTS; PRO0205; EGF10.
DR PRINTS; PRO0206; EGF10.
DR PRINTS; PRO0207; EGF10.
DR PRINTS; PRO0208; EGF10.
DR PRINTS; PRO0209; EGF10.
DR PRINTS; PRO0210; EGF10.
DR PRINTS; PRO0211; EGF10.
DR PRINTS; PRO0212; EGF10.
DR PRINTS; PRO0213; EGF10.
DR PRINTS; PRO0214; EGF10.
DR PRINTS; PRO0215; EGF10.
DR PRINTS; PRO0216; EGF10.
DR PRINTS; PRO0217; EGF10.
DR PRINTS; PRO0218; EGF10.
DR PRINTS; PRO0219; EGF10.
DR PRINTS; PRO0220; EGF10.
DR PRINTS; PRO0221; EGF10.
DR PRINTS; PRO0222; EGF10.
DR PRINTS; PRO0223; EGF10.
DR PRINTS; PRO0224; EGF10.
DR PRINTS; PRO0225; EGF10.
DR PRINTS; PRO0226; EGF10.
DR PRINTS; PRO0227; EGF10.
DR PRINTS; PRO0228; EGF10.
DR PRINTS; PRO0229; EGF10.
DR PRINTS; PRO0230; EGF10.
DR PRINTS; PRO0231; EGF10.
DR PRINTS; PRO0232; EGF10.
DR PRINTS; PRO0233; EGF10.
DR PRINTS; PRO0234; EGF10.
DR PRINTS; PRO0235; EGF10.
DR PRINTS; PRO0236; EGF10.
DR PRINTS; PRO0237; EGF10.
DR PRINTS; PRO0238; EGF10.
DR PRINTS; PRO0239; EGF10.
DR PRINTS; PRO0240; EGF10.
DR PRINTS; PRO0241; EGF10.
DR PRINTS; PRO0242; EGF10.
DR PRINTS; PRO0243; EGF10.
DR PRINTS; PRO0244; EGF10.
DR PRINTS; PRO0245; EGF10.
DR PRINTS; PRO0246; EGF10.
DR PRINTS; PRO0247; EGF10.
DR PRINTS; PRO0248; EGF10.
DR PRINTS; PRO0249; EGF10.
DR PRINTS; PRO0250; EGF10.
DR PRINTS; PRO0251; EGF10.
DR PRINTS; PRO0252; EGF10.
DR PRINTS; PRO0253; EGF10.
DR PRINTS; PRO0254; EGF10.
DR PRINTS; PRO0255; EGF10.
DR PRINTS; PRO0256; EGF10.
DR PRINTS; PRO0257; EGF10.
DR PRINTS; PRO0258; EGF10.
DR PRINTS; PRO0259; EGF10.
DR PRINTS; PRO0260; EGF10.
DR PRINTS; PRO0261; EGF10.
DR PRINTS; PRO0262; EGF10.
DR PRINTS; PRO0263; EGF10.
DR PRINTS; PRO0264; EGF10.
DR PRINTS; PRO0265; EGF10.
DR PRINTS; PRO0266; EGF10.
DR PRINTS; PRO0267; EGF10.
DR PRINTS; PRO0268; EGF10.
DR PRINTS; PRO0269; EGF10.
DR PRINTS; PRO0270; EGF10.
DR PRINTS; PRO0271; EGF10.
DR PRINTS; PRO0272; EGF10.
DR PRINTS; PRO0273; EGF10.
DR PRINTS; PRO0274; EGF10.
DR PRINTS; PRO0275; EGF10.
DR PRINTS; PRO0276; EGF10.
DR PRINTS; PRO0277; EGF10.
DR PRINTS; PRO0278; EGF10.
DR PRINTS; PRO0279; EGF10.
DR PRINTS; PRO0280; EGF10.
DR PRINTS; PRO0281; EGF10.
DR PRINTS; PRO0282; EGF10.
DR PRINTS; PRO0283; EGF10.
DR PRINTS; PRO0284; EGF10.
DR PRINTS; PRO0285; EGF10.
DR PRINTS; PRO0286; EGF10.
DR PRINTS; PRO0287; EGF10.
DR PRINTS; PRO0288; EGF10.
DR PRINTS; PRO0289; EGF10.
DR PRINTS; PRO0290; EGF10.
DR PRINTS; PRO0291; EGF10.
DR PRINTS; PRO0292; EGF10.
DR PRINTS; PRO0293; EGF10.
DR PRINTS; PRO0294; EGF10.
DR PRINTS; PRO0295; EGF10.
DR PRINTS; PRO0296; EGF10.
DR PRINTS; PRO0297; EGF10.
DR PRINTS; PRO0298; EGF10.
DR PRINTS; PRO0299; EGF10.
DR PRINTS; PRO0300; EGF10.
DR PRINTS; PRO0301; EGF10.
DR PRINTS; PRO0302; EGF10.
DR PRINTS; PRO0303; EGF10.
DR PRINTS; PRO0304; EGF10.
DR PRINTS; PRO0305; EGF10.
DR PRINTS; PRO0306; EGF10.
DR PRINTS; PRO0307; EGF10.
DR PRINTS; PRO0308; EGF10.
DR PRINTS; PRO0309; EGF10.
DR PRINTS; PRO0310; EGF10.
DR PRINTS; PRO0311; EGF10.
DR PRINTS; PRO0312; EGF10.
DR PRINTS; PRO0313; EGF10.
DR PRINTS; PRO0314; EGF10.
DR PRINTS; PRO0315; EGF10.
DR PRINTS; PRO0316; EGF10.
DR PRINTS; PRO0317; EGF10.
DR PRINTS; PRO0318; EGF10.
DR PRINTS; PRO0319; EGF10.
DR PRINTS; PRO0320; EGF10.
DR PRINTS; PRO0321; EGF10.
DR PRINTS; PRO0322; EGF10.
DR PRINTS; PRO0323; EGF10.
DR PRINTS; PRO0324; EGF10.
DR PRINTS; PRO0325; EGF10.
DR PRINTS; PRO0326; EGF10.
DR PRINTS; PRO0327; EGF10.
DR PRINTS; PRO0328; EGF10.
DR PRINTS; PRO0329; EGF10.
DR PRINTS; PRO0330; EGF10.
DR PRINTS; PRO0331; EGF10.
DR PRINTS; PRO0332; EGF10.
DR PRINTS; PRO0333; EGF10.
DR PRINTS; PRO0334; EGF10.
DR PRINTS; PRO0335; EGF10.
DR PRINTS; PRO0336; EGF10.
DR PRINTS; PRO0337; EGF10.
DR PRINTS; PRO0338; EGF10.
DR PRINTS; PRO0339; EGF10.
DR PRINTS; PRO0340; EGF10.
DR PRINTS; PRO0341; EGF10.
DR PRINTS; PRO0342; EGF10.
DR PRINTS; PRO0343; EGF10.
DR PRINTS; PRO0344; EGF10.
DR PRINTS; PRO0345; EGF10.
DR PRINTS; PRO0346; EGF10.
DR PRINTS; PRO0347; EGF10.
DR PRINTS; PRO0348; EGF10.
DR PRINTS; PRO0349; EGF10.
DR PRINTS; PRO0350; EGF10.
DR PRINTS; PRO0351; EGF10.
DR PRINTS; PRO0352; EGF10.
DR PRINTS; PRO0353; EGF10.
DR PRINTS; PRO0354; EGF10.
DR PRINTS; PRO0355; EGF10.
DR PRINTS; PRO0356; EGF10.
DR PRINTS; PRO0357; EGF10.
DR PRINTS; PRO0358; EGF10.
DR PRINTS; PRO0359; EGF10.
DR PRINTS; PRO0360; EGF10.
DR PRINTS; PRO0361; EGF10.
DR PRINTS; PRO0362; EGF10.
DR PRINTS; PRO0363; EGF10.
DR PRINTS; PRO0364; EGF10.
DR PRINTS; PRO0365; EGF10.
DR PRINTS; PRO0366; EGF10.
DR PRINTS; PRO0367; EGF10.
DR PRINTS; PRO0368; EGF10.
DR PRINTS; PRO0369; EGF10.
DR PRINTS; PRO0370; EGF10.
DR PRINTS; PRO0371; EGF10.
DR PRINTS; PRO0372; EGF10.
DR PRINTS; PRO0373; EGF10.
DR PRINTS; PRO0374; EGF10.
DR PRINTS; PRO0375; EGF10.
DR PRINTS; PRO0376; EGF10.
DR PRINTS; PRO0377; EGF10.
DR PRINTS; PRO0378; EGF10.
DR PRINTS; PRO0379; EGF10.
DR PRINTS; PRO0380; EGF10.
DR PRINTS; PRO0381; EGF10.
DR PRINTS; PRO0382; EGF10.
DR PRINTS; PRO0383; EGF10.
DR PRINTS; PRO0384; EGF10.
DR PRINTS; PRO0385; EGF10.
DR PRINTS; PRO0386; EGF10.
DR PRINTS; PRO0387; EGF10.
DR PRINTS; PRO0388; EGF10.
DR PRINTS; PRO0389; EGF10.
DR PRINTS; PRO0390; EGF10.
DR PRINTS; PRO0391; EGF10.
DR PRINTS; PRO0392; EGF10.
DR PRINTS; PRO0393; EGF10.
DR PRINTS; PRO0394; EGF10.
DR PRINTS; PRO0395; EGF10.
DR PRINTS; PRO0396; EGF10.
DR PRINTS; PRO0397; EGF10.
DR PRINTS; PRO0398; EGF10.
DR PRINTS; PRO0399; EGF10.
DR PRINTS; PRO0400; EGF10.
DR PRINTS; PRO0401; EGF10.
DR PRINTS; PRO0402; EGF10.
DR PRINTS; PRO0403; EGF10.
DR PRINTS; PRO0404; EGF10.
DR PRINTS; PRO0405; EGF10.
DR PRINTS; PRO0406; EGF10.
DR PRINTS; PRO0407; EGF10.
DR PRINTS; PRO0408; EGF10.
DR PRINTS; PRO0409; EGF10.
DR PRINTS; PRO0410; EGF10.
DR PRINTS; PRO0411; EGF10.
DR PRINTS; PRO0412; EGF10.
DR PRINTS; PRO0413; EGF10.
DR PRINTS; PRO0414; EGF10.
DR PRINTS; PRO0415; EGF10.
DR PRINTS; PRO0416; EGF10.
DR PRINTS; PRO0417; EGF10.
DR PRINTS; PRO0418; EGF10.
DR PRINTS; PRO0419; EGF10.
DR PRINTS; PRO0420; EGF10.
DR PRINTS; PRO0421; EGF10.
DR PRINTS; PRO0422; EGF10.
DR PRINTS; PRO0423; EGF10.
DR PRINTS; PRO0424; EGF10.
DR PRINTS; PRO0425; EGF10.
DR PRINTS; PRO0426; EGF10.
DR PRINTS; PRO0427; EGF10.
DR PRINTS; PRO0428; EGF10.
DR PRINTS; PRO0429; EGF10.
DR PRINTS; PRO0430; EGF10.
DR PRINTS; PRO0431; EGF10.
DR PRINTS; PRO0432; EGF10.
DR PRINTS; PRO0433; EGF10.
DR PRINTS; PRO0434; EGF10.
DR PRINTS; PRO0435; EGF10.
DR PRINTS; PRO0436; EGF10.
DR PRINTS; PRO0437; EGF10.
DR PRINTS; PRO0438; EGF10.
DR PRINTS; PRO0439; EGF10.
DR PRINTS; PRO0440; EGF10.
DR PRINTS; PRO0441; EGF10.
DR PRINTS; PRO0442; EGF10.
DR PRINTS; PRO0443; EGF10.
DR PRINTS; PRO0444; EGF10.
DR PRINTS; PRO0445; EGF10.
DR PRINTS; PRO0446; EGF10.
DR PRINTS; PRO0447; EGF10.
DR PRINTS; PRO0448; EGF10.
DR PRINTS; PRO0449; EGF10.
DR PRINTS; PRO0450; EGF10.
DR PRINTS; PRO0451; EGF10.
DR PRINTS; PRO0452; EGF10.
DR PRINTS; PRO0453; EGF10.
DR PRINTS; PRO0454; EGF10.
DR PRINTS; PRO0455; EGF10.
DR PRINTS; PRO0456; EGF10.
DR PRINTS; PRO0457; EGF10.
DR PRINTS; PRO0458; EGF10.
DR PRINTS; PRO0459; EGF10.
DR PRINTS; PRO0460; EGF10.
DR PRINTS; PRO0461; EGF10.
DR PRINTS; PRO0462; EGF10.
DR PRINTS; PRO0463; EGF10.
DR PRINTS; PRO0464; EGF10.
DR PRINTS; PRO0465; EGF10.
DR PRINTS; PRO0466; EGF10.
DR PRINTS; PRO0467; EGF10.
DR PRINTS; PRO0468; EGF10.
DR PRINTS; PRO0469; EGF10.
DR PRINTS; PRO0470; EGF10.
DR PRINTS; PRO0471; EGF10.
DR PRINTS; PRO0472; EGF10.
DR PRINTS; PRO0473; EGF10.
DR PRINTS; PRO0474; EGF10.
DR PRINTS; PRO0475; EGF10.
DR PRINTS; PRO0476; EGF10.
DR PRINTS; PRO0477; EGF10.
DR PRINTS; PRO0478; EGF10.
DR PRINTS; PRO0479; EGF10.
DR PRINTS; PRO0480; EGF10.
DR PRINTS; PRO0481; EGF10.
DR PRINTS; PRO0482; EGF10.
DR PRINTS; PRO0483; EGF10.
DR PRINTS; PRO0484; EGF10.
DR PRINTS; PRO0485; EGF10.
DR PRINTS; PRO0486; EGF10.
DR PRINTS; PRO0487; EGF10.
DR PRINTS; PRO0488; EGF10.
DR PRINTS; PRO0489; EGF10.
DR PRINTS; PRO0490; EGF10.
DR PRINTS; PRO0491; EGF10.
DR PRINTS; PRO0492; EGF10.
DR PRINTS; PRO0493; E

KW glycoprotein; EGF-like domain.
 FT NON_TER 1
 SO SEQUENCE 529 AA; 55543 MW; DAAB958FCF9ACB5A CRC64;

Query Match 12.2%; Score 114; DB 5; Length 529;
 Best Local Similarity 28.4%; Pred. No. 0.18;
 Matches 40; Conservative 22; Mismatches 55; Indels 24; Gaps 7;

OY 34 VSAEAG---ITGWTYNOLGTFVTAGADGALITGTV---ESAIVGNESRYVL---TGR 83
 DB 396 VNCSEVGECDLEGWYNECNDQIRITKTSGLDYMTAVEIIVGYAAPVYVGYASNN 455
 OY 84 YDSAPATDGSSTALGWTYVAMKNRYRNAHSATTWSGO-YVGGAEARINTOMLTSGT-TEA 141
 DB 456 YDF-----PSFGFTYVNDNGM-----STTSWTAQCHCDNEEVLTYTMINIMVDTQC 503
 OY 142 NAMKSTLVGHDTFTKVPKPSAA 162
 DB 504 DIKSKNMVGODKMTREYHSTA 524

RESULT 2
 ID 006353 PRELIMINARY; PRT; 595 AA.
 AC 006353;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE FLAGELLIN.
 GN FLIC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BI 316-42 / SEROTYPE 09:K9:H12;
 RX MEDLINE=93374833; PubMed=8366026;
 RA Schoenhalz G., Whitfield C.;
 RT "Comparative analysis of flagellin sequences from Escherichia coli strains possessing serologically distinct flagellar filaments with a shared complex surface pattern."
 RT J. Bacteriol. 175:5395-5402(1993).
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
 CC EMBL: I07389; AAA23799.1; -.
 DR INTERPRO: IPR001029; -.
 DR INTERPRO: IPR001492; -.
 DR PFAM: PF00669; Flagellin_N; 1.
 DR PFAM: PF00700; Flagellin_C; 1.
 DR PRINTS: PR00207; FLAGELLIN.
 DR PRODOM: PD000316; -; 1.
 KW Flagella.
 SO SEQUENCE 595 AA; 61044 MW; 88510CD561EF25F9 CRC64;

Query Match 12.0%; Score 112; DB 2; Length 595;
 Best Local Similarity 26.1%; Pred. No. 0.29;
 Matches 52; Conservative 30; Mismatches 87; Indels 30; Gaps 9;

OY 7 AATAVSLTTSITASADPSKSKAQSAAEAGITGTWYNOLGST---FIYTAGADGAL 63
 DB 343 ASVTMGCTTYNFRGTGADAD-AATANAGVSFTDTASKETVLANVATAKQKAVAAADGDTSA 401
 OY 64 TGTYESAV-----GNAESRYVLGRYDSAPAT---DGSSTALG-RTVAMK--- 104
 DB 402 TTYTKSGVOTYQAVFAAGGTASAKYADKADVSNATATYTDADGEMTTGTTTKYSIDA 461
 OY 105 NMYRNAHSATTWSGOYVG--GAEARINTOWLTS-GTTANAMKSTLVGHG---TFTKV 157
 DB 462 NNGKTVVDSGTGTGAKPRVGAEEYVYANGTLTTDTATSGTVDKPLKALDEAISISIDKF 521

OY 158 KPSAASIDAAKKAGVNNGN 176
 DB 522 RPSLGATONRLDSAVTNLN 540

RESULT 3
 ID 09KX38 PRELIMINARY; PRT; 892 AA.
 AC 09KX38;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE USPA1.
 GN USPA1.
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
 OX NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O46E;
 RX MEDLINE=20138164; PubMed=10671460;
 RA Lafontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr., Hansen E.J.;
 RT "The USPA1 Protein and a Second Type of UsPA2 Protein Mediate Adherence of Moraxella catarrhalis to Human Epithelial Cells in vitro."
 RT J. Bacteriol. 182:1364-1373(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O46E;
 RA Maciver I., Latimer J.L., Cope L.D., Thomas S.E., Hansen E.J.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U61725; AAF36416.1; -.
 SO SEQUENCE 892 AA; 93358 MW; 9D71A369672F44C7 CRC64;

Query Match 11.8%; Score 110.5; DB 2; Length 892;
 Best Local Similarity 21.7%; Pred. No. 0.59;
 Matches 43; Conservative 26; Mismatches 86; Indels 43; Gaps 6;

OY 17 SITASADPSKSKAQSAAEAGITGTWYNOLGSTFIYTAGADGALT--GTYESAVGNA 74
 DB 101 SYVGGSSSWTAGKESKSTIGGDTNDANGYTSITIGGYSRALGDSSTIGGTYNQTG-- 158
 OY 75 ESTRYVTGTGDSAPATDGSCTALGWTVAMKNRY--NAHSATTWSGOYVGAEARINTQ 131
 DB 159 EKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSHNOATGEGSPFAGVENKANAN 218
 OY 132 -----WLTSGTTEANAK-STLVGHDTFTKYK 158
 DB 219 NAVAIGKNTTIDGDSVAIGSNNTIDSGKQNVFTLGSSSTNTTNAQSGSVLLGHNTAGK-- 276
 OY 159 PSAASIDAAKKAGVNNGN 176
 DB 277 -KATAVSSAKVNGTLIGN 293

RESULT 4
 ID 005164 PRELIMINARY; PRT; 1001 AA.
 AC 005164;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-AUG-1998 (TREMblrel. 07, Last annotation update)
 DE AOB567, AOF1001, AOE110, AOE264 AND AOE130 GENES.
 GN AOF1001.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FY1679;

RA Gamo F.J., Lafuente M.J., Casamayor A., Aldea M., Casas C., Arto J.,
 RA Herrero E., Gancedo C.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X89715; CA61860.1;
 SO SEQUENCE 1001 AA; 97673 MW; F11BC4522DCFB04F CRC64;

Query Match 11.3%; Score 105.5; DB 3; Length 1001;
 Best Local Similarity 24.7%; Pred. No. 1.7;
 Matches 42; Conservative 33; Mismatches 64; Indels 31; Gaps 5;

OY 6 VAAIAVSLTIVSTIASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAG----ADG 61
 DB 63 VSSSLRELTSSSTREVSSTAPSTSS---SEVSSITSSGSSVSGSSSITSSGSSVSSSS 118
 OY 62 ALGTGESAVGNAESRYVLGTRDAPATDGGTALGWTYAMKNNRNHSAATTWGGYV 121
 DB 119 SATSSSSSASGSSSAT-----EGSSSVSGSSSTSI-----TSGSSSATESGSSV 161
 OY 122 GGAEARINTQWLTSCTTEANAMKSTLVGHDFTKVYKPSAASIDAKKAG 171
 DB 162 SSSSTSI-----TSGSSSATESGSSVSGSTSATSSGSSSASGSSSATESG 205

RESULT 5
 ID 09SOT7 PRELIMINARY: PRT: 595 AA.
 AC 09SOT7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE FLAGELLIN.
 GN FLIC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BI 316-42;
 RA Ohnishi K., Ishiooka K., Matsuba T., Hareyama S.,
 RT "Cloning of H antigen genes in E.coli serotypes and expression in
 RT E.coli K-12."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB028475; BAA85084.1;
 DR INTERPRO: IPR001029;
 DR INTERPRO: IPR001492;
 DR PFAM: PF00669; Flagellin_N; 1.
 DR PFAM: PF00700; Flagellin_C; 1.
 DR PRINTS: PR00207; FLAGELLIN.
 SO SEQUENCE 595 AA; 61020 MW; 8B9DC3D6AC78C427 CRC64;

Query Match 11.1%; Score 104; DB 2; Length 595;
 Best Local Similarity 25.6%; Pred. No. 1.2;
 Matches 51; Conservative 30; Mismatches 88; Indels 30; Gaps 9;

OY 7 AAIAVSLTIVSTIASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADGAL 63
 DB 343 ASYTMGCTTYNFKTGADAD-AATANAGVSTPDASKETVINKVATKQGRVAAADGDTSA 401
 OY 64 TGTYESAV-----GNAESRYVLGRYDSAPAT-----DGSSTALG-WTVAMK--- 104
 DB 402 TTYKSGVQYQAVFAAGDGTASAKYADNDVSNATVYTDADGEMTTIGSYTTKYSIDA 461
 OY 105 NNYRNHSAATTWGGYV--GAEARINTQWLTS-GTTEANAMKSTLVGHG----FTTKV 157
 DB 462 NNGKVTYVDSGTGKYPKGAEVYVSANGTLTTDATSEGTVTKDPLKALDEAISIDKF 521
 OY 158 KPSAASIDAKKAGVNNGN 176
 DB 522 RSSLGAIONRLDSAVTNLN 540

RESULT 6
 ID 006351 PRELIMINARY: PRT: 595 AA.
 AC 006351;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE FLAGELLIN.
 GN FLIC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SU 1242 / SEROTYPE 01:K2:H1;
 RX MEDLINE=93374833; PubMed=8366026;
 RA Schoenheits G., Whitfield C.;
 RT "Comparative analysis of flagellin sequences from Escherichia coli
 RT strains possessing serologically distinct flagellar filaments with a
 RT shared complex surface pattern."
 RL J. Bacteriol. 175:5395-5402(1993).
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
 CC -1- SIMILARITY: NO OTHER BACTERIAL FLAGELLINS.
 DR EMBL: L07387; AAA23797.1;
 DR INTERPRO: IPR001029;
 DR INTERPRO: IPR001492;
 DR PFAM: PF00669; Flagellin_N; 1.
 DR PFAM: PF00700; Flagellin_C; 1.
 DR PRINTS: PR00207; FLAGELLIN.
 DR PRODOM: PD000316; 1.
 KW Flagella.
 SO SEQUENCE 595 AA; 60923 MW; 9AAFB2B6884607A CRC64;

Query Match 10.8%; Score 101; DB 2; Length 595;
 Best Local Similarity 25.6%; Pred. No. 2.1;
 Matches 51; Conservative 29; Mismatches 89; Indels 30; Gaps 9;

OY 7 AAIAVSLTIVSTIASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADGAL 63
 DB 343 ASYTMGCTTYNFKTGADAG-AATANAGVSTPDASKETVINKVATKQGRVAAADGDTSA 401
 OY 64 TGTYESAV-----GNAESRYVLGRYDSAPAT-----DGSSTALG-WTVAMK--- 104
 DB 402 TTYKSGVQYQAVFAAGDGTASAKYADNDVSNATVYTDADGEMTTIGSYTTKYSIDA 461
 OY 105 NNYRNHSAATTWGGYV--GAEARINTQWLTS-GTTEANAMKSTLVGHG----FTTKV 157
 DB 462 NNGKVTYVDSGTGKYPKGAEVYVSANGTLTTDATSEGTVTKDPLKALDEAISIDKF 521
 OY 158 KPSAASIDAKKAGVNNGN 176
 DB 522 RSSLGAIONRLDSAVTNLN 540

RESULT 7
 ID 09RF12 PRELIMINARY: PRT: 1026 AA.
 AC 09RF12;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE S-LAYER PROTEIN.
 GN R5AA.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=76;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-JS3001;
 RA Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;
 RT "The Secretion Signal of C. crescentus S-layer Protein is Located in
 the C-terminal 82 Amino Acids of the Molecule.";
 RL Submitted (Oct-1999) to the EMBL/Genbank/DBU databases.
 DR EMBL; AF193063; AAF19365.1; -
 DR INTERPRO: IPR001343; -
 DR INTERPRO: IPR001952; -
 DR PEAM: PR00353; hemolysincbind; 1.
 DR PRINTS; PR00313; CABDNCRPT.
 DR PROSITE; PS00123; ALKALINE_PHOSPHATASE; UNKNOWN_1.
 DR SEQUENCE 1026 AA; 98132 MW; AD7DB818D7C528AC CRC64;

Query Match 10.7%; Score 100.5; DB 2; Length 1026;
 Best Local Similarity 28.8%; Pred. No. 4.2;
 Matches 55; Conservative 18; Mismatches 85; Indels 33; Gaps 7;

OY 6 VAAIAVSLTVSITASADPSKDSKAQV-----SAAEAGITGWYNOLGSTFIYTAGADG 61
 DB 299 VQAAATVLTPTGVTIGTIEFMNTSGAATITLNTSSGVTGLTALNTTSGAAQTVTAGAGQ 358
 OY 62 ALTGYESAV-----GNASRYVLTRYDSAPATDGSCTALGWTVMKNRYNAHSAT 114
 DB 359 NLTATTAQAQANNVAVDGGANTVASTG-VTSGTTVGANSAASGTV---SVVANSST 413
 OY 115 TWSGQY-VGGAERINTOMLTSGTTEANAKSTLYGHD-----TFTKYKPSAS 164
 DB 414 TTTGTAIVTGTGTR-----VTVAQTAGNANVTTLTQADVTYTGNSSTYAVTYQTAAAT 466
 OY 165 DAAKAGVNNNG 175
 DB 467 AGATVAGRVNG 477

RESULT 8
 OY9S4M2 PRELIMINARY; PRT; 560 AA.
 ID OY9S4M2
 AC OY9S4M2
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLIC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E7A/68;
 RX MEDLINE-99084952; PubMed-9864325;
 RA Reid S.D., Selander R.K., Whittam T.S.;
 RT "Sequence diversity of flagellin (flc) alleles in pathogenic
 RT Escherichia coli.";
 RL J. Bacteriol. 181:153-160(1999).
 DR EMBL; AF128954; AAD28525.2; -
 DR INTERPRO: IPR001023; -
 DR INTERPRO: IPR001029; -
 DR INTERPRO: IPR001029; -
 DR INTERPRO: IPR001492; -
 DR PEAM; PF00669; Flagellin_N; 1.
 DR PEAM; PF00700; Flagellin_C; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR PRINTS; PR00301; HEATSHOCK70.
 DR PRINTS; PR00455; HTHETTR.
 FT NON_TER 1
 FT SEQUENCE 560 AA; 57262 MW; 97DF005C9E0ABB71 CRC64;

Query Match 10.7%; Score 100; DB 2; Length 560;
 Best Local Similarity 25.1%; Pred. No. 2.4;

Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;
 OY 7 AATAVSLTVSITASADPSKDSKAQVSAEAGITGWYNOLGST---FIYTAGADGAL 63
 DB 324 ASVTMGSTTYNFKTGDAD-AAATANAGVSFTDPAKSEYVLNKAATAKQKRAAADDGTS 382
 OY 64 TGTYESAV-----GNASRYVLTRYDSAPAT-----DGSCTALG-WTVAMK--- 104
 DB 383 TITTKSGVQTYQAVFAAGDGTASAKYADKADVSNATRTYTDADGENTTIGSYTTKTSIDA 442
 OY 105 NNYRNAHSATTWGQYVG--GAERINTOMLTS-GTTEANAKSTLYGHD---TFTKY 157
 DB 443 NNGKVTYVDSGTGKTAAPKVGAEVYSANGTLTTDATSEGTYRKDPLKALDEAIISSIDKF 502
 OY 158 KPSASITDAAKKAGVNNGN 176
 DB 503 RSSLGAIQNRLDASVTNLN 521

RESULT 9
 OY9S4M4 PRELIMINARY; PRT; 565 AA.
 ID OY9S4M4
 AC OY9S4M4
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLIC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DEC 6A;
 RX MEDLINE-99084952; PubMed-9864325;
 RA Reid S.D., Selander R.K., Whittam T.S.;
 RT "Sequence diversity of flagellin (flc) alleles in pathogenic
 RT Escherichia coli.";
 RL J. Bacteriol. 181:153-160(1999).
 DR EMBL; AF128952; AAD28523.2; -
 DR INTERPRO: IPR001023; -
 DR INTERPRO: IPR001029; -
 DR INTERPRO: IPR001492; -
 DR INTERPRO: IPR001647; -
 DR PEAM; PF00669; Flagellin_N; 1.
 DR PEAM; PF00700; Flagellin_C; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR PRINTS; PR00301; HEATSHOCK70.
 DR PRINTS; PR00455; HTHETTR.
 FT NON_TER 1
 FT SEQUENCE 565 AA; 57857 MW; 6E9EEA1A6F72DC29 CRC64;

Query Match 10.7%; Score 100; DB 2; Length 565;
 Best Local Similarity 25.1%; Pred. No. 2.4;
 Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;

OY 7 AATAVSLTVSITASADPSKDSKAQVSAEAGITGWYNOLGST---FIYTAGADGAL 63
 DB 328 ASVTMGSTTYNFKTGDAD-AAATANAGVSFTDPAKSEYVLNKAATAKQKRAAADDGTS 386
 OY 64 TGTYESAV-----GNASRYVLTRYDSAPAT-----DGSCTALG-WTVAMK--- 104
 DB 387 TITTKSGVQTYQAVFAAGDGTASAKYADKADVSNATRTYTDADGENTTIGSYTTKTSIDA 446
 OY 105 NNYRNAHSATTWGQYVG--GAERINTOMLTS-GTTEANAKSTLYGHD---TFTKY 157
 DB 447 NNGKVTYVDSGTGKTAAPKVGAEVYSANGTLTTDATSEGTYRKDPLKALDEAIISSIDKF 506
 OY 158 KPSASITDAAKKAGVNNGN 176

Db 507 RSSIGAIONRLDSAVTNLN 525

RESULT 10

Q48028 PRELIMINARY; PRT; 1477 AA.

AC Q48028; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)

DE ADHESIN.

GN HMM2A.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus

OX NCBI_Taxid=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=12;

RX MEDLINE=92192797; PubMed=1548058;

RA Barenkamp S.J., Leininger E.;

RT "Cloning, expression, and DNA sequence analysis of genes encoding nonlysozyme Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis."

RT pertussis."

RL Infect. Immun. 60:1302-1313(1992).

DR EMBL: U08875; AAA20524.1; -.

DR INTERPRO: IPR000169; -.

DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.

DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.

SO SEQUENCE 1477 AA; 154472 MW; B057C23F1AD24B0E CRC64;

Query Match 10.6%; Score 99.5; DB 2; Length 1477;
Best Local Similarity 26.2%; Pred. No. 7.6;
Matches 50; Conservative 26; Mismatches 88; Indels 27; Gaps 6;

QY 11 VSLTTSITTSASADPSKDSKAQVSAAEAGI---GTWYNQL-GSTFIVTAGDALTG 66
DB 1177 ISGNTVSASATGDLTTKSGSKTEAKSGEANTVSATGTIGTISGNTVNAVANA-GDLTVG 1235
QY 67 YESAVGNAESRYVLTGRYDSAPATDGS-----TALGWTVAMKNRYNAHSATTW 116
DB 1236 NGAFINTEGATITATNGNLTTEAGSSITSTKQVDLLAONGIASINAAVTLWTG 1295
QY 117 SCGVGGAEARINTQWLLTSGTTEANAMKSTLVGHDTFTKKVPSA-----ASIDAARKA 170
DB 1296 TLTIVAGSDIK-----ATSGTLVINAKDAKLNGDASGDSREVNANVANSAGSVTAATSS 1349
QY 171 GVNNGNPLDAV 181
DB 1350 SVNITGDLNTV 1360

RESULT 11

054356 PRELIMINARY; PRT; 832 AA.

AC 054356; 01-JUN-1998 (TREMblrel. 06, Created)

DT 01-JUN-1998 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)

DE HIGH MOLECULAR WEIGHT OUTER MEMBRANE PROTEIN.

GN USPA1.

OS Moraxella catarrhalis.

OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.

OX NCBI_Taxid=480;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=035E;

RX MEDLINE=98013056; PubMed=9353007;

RA Abbl C., MacIver I., Latimer J.L., Cope L.D., Stevens M.K.,

Thomas S.E., McCracken G.H. Jr., Hansen E.J.;

RT "A protective epitope of Moraxella catarrhalis is encoded by two different genes."

RL Infect. Immun. 65:4367-4377(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=035E;

RA Hansen E.J.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: U57551; AAB96359.2; -.

DR INTERPRO: IPR000519; -.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.

DR PROSITE: PS00402; BPD_TRANSF. INN MEMBR; UNKNOWN_1.

SO SEQUENCE 832 AA; 88292 MW; 5C47F4F273350F4B CRC64;

Query Match 10.5%; Score 98.5; DB 2; Length 832;
Best Local Similarity 22.0%; Pred. No. 4.8;
Matches 54; Conservative 34; Mismatches 71; Indels 87; Gaps 11;

QY 2 KRIYVAIANSLTIVTTSASADPSKDSKAQV----- 34
DB 28 KAVGSLIIVGALGMATTAQAATNSKGTGAHGVNNNEAPGDSYFISGGYKKAEGRY 87
QY 35 SAAGAG-----ITGWNQL-GSTFIVTAGDALTGYESAVGNAESRYVLTG 82
DB 88 SAIGGGLEPKATNEYSTIVGGYKKAEGRYSTIGGSNNEATNEYSTIVGGDDNK--ATG 145
QY 83 RYDSAPATDGSSTALG--WTVA--WKNNYRNAHSATWISGOVGAEARINTQ----- 131
DB 146 RYSTIGGD--NNTAFEGRYSTIVAGSKN-----QATGTGSAAGVENQANENAVAVGK 198
QY 132 -----WLTSGTTEANAMKSTLVGHDTFTKKVPSAASIDAA 167
DB 199 NIIEGNSVAIGSENVKTEHKNFILGSGTGYTS--NSVLLGNETAGK--QATTVKNA 254
QY 168 KKAQVN 173
DB 255 EVGGLS 260

RESULT 12

09S001 PRELIMINARY; PRT; 595 AA.

AC 09S001; 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE FLAGELLIN.

GN FLIC.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_Taxid=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SU 1242;

RA Omishi K., Ishioke K., Matsuba T., Harayama S.;

RT "Cloning of H antigen genes in E.coli H serotypes and expression in E.coli K-12."

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB028471; BAA65080.1; -.

DR INTERPRO: IPR001029; -.

DR PROSITE: PS00669; Flagellin_N; 1.

DR PROSITE: PS00700; Flagellin_C; 1.

DR PRINTS: PR00207; FLAGELLIN.

SO SEQUENCE 595 AA; 60937 MW; 9AAC42EF6884607D CRC64;

Query Match 10.5%; Score 98; DB 2; Length 595;
Best Local Similarity 25.1%; Pred. No. 3.7;
Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;

KM Hydrolase.
SQ SEQUENCE 507 AA; 52681 MW; 5EAD34A9874C6A73 CRC64;

Query Match	10.1%	Score 94.5	DB 2	Length 507
Best Local Similarity	27.6%	Pred No. 5.8		
Matches 43	Conservative 16	Mismatches 52	Indels 45	Gaps 10

Best Local Similarity 27.6%; Pred. No. 5.8;
 Matched 43; Corresponding 16; Mismatched

Matches	43;	Conservative	16;	Mismatches	52;	Indels	45;	Gaps	10;
---------	-----	--------------	-----	------------	-----	--------	-----	------	-----

```

0Y 10 AVSLTUVV--TTASASADPS-----KOSKO-----VSAEAGITGWYNQLOSTFIYTA 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35S SL$ATPAGCTVITAGASATTTVTKAVKSGTAQTVOLGASGVPAQVTA$F-----SPASVTA 409A
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 58 GADGALT-----GTYESAV-----GNAESRYVLGTRHSDAPATDGS--TALGW 99
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 GGOSTLILATTGGAAVSGSYSLTITGCTGPGBSHSTYALT-----VTGANGNOCTAVPW 462A
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 100 T--VAMKNNVNRNASHATWSGO--VVGGAEARINTQW 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 NSGALITYGGQOV$HEGHTWKKKKMTTGEPEPTQGW 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 355 SLSATPASGTVTAGASATTTVKTAVKSGTAQTVOLGASGVPAGVTASF-----SPASVTA 409

QY 58 GADGALT-----CTYESAV-----GNAESRYVLTGRYDSAPATDGS--TALGW 99

Db 410 GGSTLTATTGAVSGTYSITVTGTGSPSGSHSTTYALT-----VTGGNGNOCTAVPW 462

OY 100 T--VAMKNNYRNAHSATTWSCG-YVGGAEARINTQW 132

Db 463 NSGAIYTGQOVSHGHTWKAKWTTTGEEPTTGQW 498

Search completed: April 6, 2001, 00:18:32
Job time: 151 sec

Job time: 151 sec

THIS PAGE BLANK (USPTO)

OM of: US-09-589-870-2 to: EST:* out_format: pfs
Date: Apr 6, 2001 12:34 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+g2n.model -DEV=xlp
-O=/usr1_1/USPTO.spool/us09589870/runat_02042001_145218_15070/app_query.fasta_1.242
-DB=EST -OEPF=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosome62 -TRANS=human40.cdi
-LIST=45 -DOCAIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09589870@cgn1.1.2225 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-589-870-2
Query Length: 183
Database: EST:*
Database sequences: 7991742
Database length: -791223438
Search time (sec): 891.830000

score_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_est47:AM632946	+	132.50	226.75	0.0013	495	AM632946 bl02a11.x1 Blackshear/
gb_est14:AF122254	+	107.00	195.75	0.0683	546	AF122254 AF122254 Strongylocent
gb_gss26:CNS022YM	+	99.00	173.55	1.18	1038	AF121461 Tetradon nigroviridis
gb_est24:AI770339	-	95.50	173.25	1.22	560	AI770339 606061B11.x1 606 - Bar
gb_est79:C05809	+	94.50	172.39	1.22	506	C05809 C05809 Human pancreatic
gb_est36:AV432953	+	94.50	171.82	1.47	534	AV432953 AV432953 Porphyra yagc
gb_est47:AM672938	-	94.00	172.06	1.43	476	AM672938 WS1_12.H09.bl.A002 wat
gb_est70:BE35782	-	94.00	172.04	1.43	477	BE35782 DG1_11.A01.bl.A002 Dat
gb_est70:BE357728	-	94.00	172.04	1.43	477	BE357728 DG1_22.C07.bl.A002 Dat
gb_est70:BE357007	-	94.00	171.69	1.43	493	BE357007 DG1_146.D05.bl.A002 Dat
gb_est70:BE359336	-	94.00	171.32	1.57	511	BE359336 DG1_146.D05.bl.A002 Dat
gb_est70:BE361106	-	94.00	171.05	1.62	524	BE361106 DG1_69.F08.bl.A002 Dat
gb_est70:BE367620	-	94.00	170.78	1.68	538	BE367620 DG1_21.C08.bl.A002 Dat
gb_est70:BE357934	+	93.00	170.39	1.77	538	BE357934 DG1_23.D10.bl.A002 Dat
gb_est71:BE436234	+	93.00	169.95	1.87	613	BE436234 EST407332 tomato break
gb_gss9:AO638017	-	93.00	167.47	2.57	613	AO638017 927P1-7E4.TP 927P1 Try
gb_gss2:AO163050	-	93.00	164.89	3.58	784	AO163050 mgxb0021N13r CUGI Rice
gb_est21:AI517559	-	91.50	164.57	3.72	867	AI517559 GH28575.5prlme GH Dros
gb_est77:BE901886	+	91.50	160.94	5.94	867	BE901886 601675114F1 NIH.MGC.21
gb_est73:BE614887	+	91.50	160.65	6.16	891	BE614887 60179195F1 NIH.MGC.39
gb_est70:BE359657	-	90.50	165.95	3.12	447	BE359657 DG1_56.G03.bl.A002 Dat
gb_gss26:CNS020H0	+	90.50	158.62	7.99	899	AI197960 Tetradon nigroviridis
gb_est70:BE357417	+	90.00	161.62	5.44	616	BE357417 DG1_15.H08.bl.A002 Dat
gb_gss27:CNS04AIV	+	90.00	157.67	9.03	898	AI282028 Tetradon nigroviridis
gb_est46:AM583635	+	89.50	161.60	5.45	563	AM583635 la02c06.y1 Human Pancre
gb_est71:BE358345	-	89.00	161.31	5.31	517	BE358345 EST413637 tomato devel
gb_est41:BE358444	-	89.00	160.65	5.66	528	BE358444 EST413736 tomato devel
gb_est41:AM220895	-	89.00	159.30	7.32	639	AM220895 EST297364 tomato fruit
gb_est51:AM2930067	-	88.50	159.16	7.46	591	AM2930067 EST40524 tomato fruit
gb_est43:AM2919131	+	88.50	159.16	7.46	591	AM279131 sg06e08.y1 Gm-cl019 GI
gb_est43:AM2919131	+	88.50	159.16	7.46	591	AM279131 sg06e08.y1 Gm-cl019 GI
gb_est46:AM583588	+	88.50	158.12	7.49	593	AM583588 la01e09.y1 Human Pancre
gb_est41:AM201674	+	88.50	158.47	8.15	631	AM201674 st05h11.y1 Gm-cl027 GI
gb_est43:AM2919131	+	88.50	158.30	10.76	776	AM2919131 sg06e08.y1 Gm-cl019 GI
gb_est43:AM2919131	+	88.50	158.30	10.76	776	AM2919131 sg06e08.y1 Gm-cl019 GI
gb_est68:BE213859	+	88.50	154.66	13.28	907	BE213859 HV.CEBO001H14f Hordeum
gb_est11:AA759686	+	88.00	160.89	5.98	457	AA759686 cb19d05.s1 NCI CGAP K1
gb_est81:H00427	-	87.50	161.67	5.41	387	H00427 y123g11.r1 Soares placen
gb_est10:AA675092	+	87.50	160.15	6.57	447	AA675092 vq03d06.r1 Stratagene
gb_est71:BE427718	+	87.50	156.23	10.87	650	BE427718 PSR7380 ITEC PSR Wheat
gb_est46:AM583814	+	87.50	154.92	12.84	736	AM583814 la04h11.y1 Human Pancre
gb_est70:BE355096	+	87.00	156.94	9.92	554	BE355096 DG1_113.B03.bl.A002 Da

gb_est73:BE354518 + 87.00 156.90 9.97 556 | BE354518 P11_34.C06.bl.A002
gb_gss20:AZ350886 - 87.00 155.79 11.49 618 | AZ350886 IM0088E07R Mouse 10
gb_est71:BE427678 - 87.00 155.26 12.30 650 | BE427678 PSR7334 ITEC PSR Wh
gb_est66:BE034472 - 87.00 153.48 15.45 770 | BE034472 MH05E02 MH Mesembry

seq_name: gb_est47:AM632946

seq_documentation_block:

LOCUS AM632946 495 bp mRNA EST 03-APR-2000
DEFINITION bl02a11.x1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0002A11 3', mRNA sequence.
ACCESSION AM632946
VERSION AM632946.1 GI:7390026
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 495)
REFERENCE
AUTHORS Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Straffe,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
J.W., Bonaldo,M.F. and Soares,M.B.
The NIHs Xenopus Maternal EST Project
Unpublished (2000)
CONTACT: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
Phone: 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com

DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCAGT
BACKWARD: CAGGAACGATGACG
Plate: 0002 row: A column: 11
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers
1..495
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0002A11"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/cissue-type="unfertilized egg"
/cell-type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH108"

/note="Vector: pT73-Pac; Site1: EcoRI; Site2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adaptors, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT73-Pac vector.
The library contained approximately 7.2 x 10⁵
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT 120 a 138 c 140 t
ORIGIN
alignment_scores:

[illegible]

```

417 GCCATCGGT...GACGTCCGCTCCGCGCTGTCCTCCCAAGCGCCGAGGAC 463
      |||::| |||::| ::::|::| || ||| ::|
164 e...AspAlaAlaLysLysAlaGlyValAlaAsnAngLysnProLeu 178
      |||::|::| |||::|::| ::::|::|
464 TGGTGCAGCGATCCGCGCTGCTGCTCAGGACGACATGGACCCCTTC 509
      |||::|::| |||::|::| ::::|::|

seq_name: gb_est47:AM677938

seq_documentation_block:
LOCUS      AM677938          476 bp      mRNA          EST          19-JUL-2000
DEFINITION WS1.12_H09.bl_A002 Water-stressed 1 (ws1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AM677938
VERSION     AM677938.1  GI:7551654
KEYWORDS    EST.
SOURCE      sorghum.
ORGANISM    Sorghum bicolor
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE   1 (bases 1 to 476)
AUTHORS     Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
            An EST database from Sorghum: water-stressed plants
            Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: emmpratt@uga.edu
            Sequences have been trimmed to exclude polyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 405
            POLYA-No.

FEATURES             location/qualifiers
     source           1..476
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_1lb="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from polyA-RNA in the
                     cloning vector lambda Zap II. Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      102 a      141 c      173 g      60 t
ORIGIN
1  ATAAATSerLeuThrValSerIleThrAlaSerAlaSerAlaSpr 26
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
458 TCGCTTCTTACCTCC...TCCGCTCTGCTGCGCGCGGCGCTCCGCGC 412
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
26  oSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGlyIleThrG 43
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
411 CCGCTGCTGCTCCCTGCTCTGCTCTGCTGCTGCGCGGCGCTCAGCACGG 362
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
43  lyThrTrpIyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
361 CCGTGCAGCTGACCTTCGTGCTGCTGCTGCGACGAGACGACTTCGCCG 312

```

```

60 AspglyalaleuthrGlyThrTyrgluserAlaValgAsnaAlaGluse 76
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 TCGGCTCTCTTA...GGACCTCGGTCCGCCGCCGAGCTGCTCTC 265
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 rArgTyValleuthrGlyArgTyraSpserAlaProAlaThrAspGly 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
264 GCGGCGCGGGCTCTCTTCGACACCGGAGCGAGCCCTCCTTGCGTG 215
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 ..SerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnaSntTyra 108
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 TTTCCTCAACGCGCGCGCTGACAGCC.....TCTCCACCTAC 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 AsnaAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
173 ACCGAGGCTCTCTCGACGGCGACGAGTCGGACCTGAGCTCGCGGT 124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
120 rValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrpLeu 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 GGCCTATTGGTGCA...GCTAGGCTG.....CTGTGGTTG 93

```

seq_name: gb_est70:BE355782

seq_documentation_block: 477 bp mRNA EST 20-JUL-2000
 LOCUS BE355782 477 bp mRNA EST 20-JUL-2000
 DEFINITION DGL_11_A01.b1_A002 Dark Grown 1 (DGL) Sorghum bicolor cDNA, mRNA

ACCESSION BE355782
 VERSION BE355782
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 477)
 Magnoliophyta: Liliopsida: Poales: Poaceae; Sorghum.
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 /L.H.

TITLE An EST database from Sorghum: dark-grown seedlings
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM

Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV

High quality sequence stop: 385
 POLYA=No.

FEATURES
 source 1..477 Location/Qualifiers

/organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGL)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector:
 Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
 made from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 96 a 146 c 174 g 61 t
 ORIGIN

alignment_scores:

Quality: 94.00 Length: 130
 Ratio: 1.106 Gaps: 7
 Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:
 US-09-589-870-2 x BE355782/rev ..

Align seq 1/1 to reverse of: BE355782 from: 1 to: 477

```

10 AlAValSerLeuThrThrValSerIleThrAlaSerAlaAspPr 26
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
437 TCTGCTTCTTGACCTCC...TCCGTCTCTGCTCCGCCGCGCTCGGC 391
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
26 oSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGlyIleThr 43
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 GCGCTCTGCTCTTGTTGGTCTGTCTGCTCGCTCGCGCGCTACACGG 341
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 1yThrTrpTyraGlnLeuGlySerThrPheIleValThrAlaGlyAla 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 CCTGACCTCGACCTCCTTGGTGTCTGACGTCTCGGACGAGACGCTCGCC 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 AspglyalaleuthrGlyThrTyrgluserAlaValgAsnaAlaGluse 76
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 TCCGCTCTCTTA...GGACCTCGGTCTCCGCCGCCGAGCTGCTGTC 244
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 rArgTyValleuthrGlyArgTyraSpserAlaProAlaThrAspGly 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
243 GCGGCGCGGGCTCTCTCGACACCGGAGCGAGCCCTCCTTGCGTG 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 ..SerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnaSntTyra 108
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 TTTCCTCAACGCGCGCGCTGACAGCC.....TCTCCACCTAC 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 AsnaAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 ACCGAGGCTCTCTCGACGGCGACGAGTCGGACCTGAGCTCGCGGT 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
120 rValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrpLeu 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 GGCCTATTGGTGCA...GCTAGGCTG.....CTGTGGTTG 72

```

seq_name: gb_est70:BE357728

seq_documentation_block: 477 bp mRNA EST 20-JUL-2000
 LOCUS BE357728 477 bp mRNA EST 20-JUL-2000
 DEFINITION DGL_22_C07.b1_A002 Dark Grown 1 (DGL) Sorghum bicolor cDNA, mRNA

ACCESSION BE357728
 VERSION BE357728
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 477)
 Magnoliophyta: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
 Magnoliophyta: Liliopsida: Poales: Poaceae; Sorghum.
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 /L.H.

TITLE An EST database from Sorghum: dark-grown seedlings
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM

Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV

High quality sequence stop: 467
 POLYA=No.

FEATURES
 source 1..477 Location/Qualifiers

/organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGL)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector:
 Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
 made from poly-A RNA in the cloning vector lambda Zap II.

BASE COUNT 98 a 144 c 174 g 61 t
 ORIGIN

alignment_scores:

Quality: 94.00 Length: 130
 Ratio: 1.106 Gaps: 7
 Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:

US-09-589-870-2 x BE357728/rev ..

Align seg 1/1 to reverse of: BE357728 from: 1 to: 477

```

10 AlValSerLeuThrValSerIleThrAlaSerAlaAspPr 26
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
452 TCTGCTCTGACCTCC...TCCGCTCTGCTGCGCGGCTCGGC 406
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
26 OserLyAspSerLySAlaGlnValSerAlaAlaGlyIleThrg 43
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
405 CGGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 356
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
43 LThrTPTrpYrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
355 CCTGACCTGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
60 AspGlyAlaLeuThrGlyThrTyrgLuserAlaValGlyAsnAlaGluse 76
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
305 TCCGCTCTGCTTA...GCGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 259
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
76 rArgTyValLeuThrGlyArgTyraSpSerAlaProAlaThrAspGly. 92
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
258 GGGCGGGGGCTCTCTGACACCGACGAGCGCTCTCTGCTGCTG 209
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
93 ..SerGlyThrAlaLeuGlyThrPThrValAlaTrpLysAsnAsnTyraArg 108
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
208 TTTCCTCAACGGCGCGGCGGCGACGCC.....TCTCCACATACC 168
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
109 AsnAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
167 ACCGAGCCTCTCTGACGCGGACGCGGCGGACCTGACCTGCGCGGT 118
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
120 rValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 133
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
117 GGCCTATTGGTCA...GCTAGGGTG.....CTGTGGTTG 87

```

seq_name: gb_est70:BE357007

seq_documentation_block:

LOCUS BE357007 493 bp mRNA EST 20-JUL-2000
 DEFINITION DGI_146.D05.b1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
 sequence.

ACCESSION BE357007
 VERSION BE357007.1 GI:9298564

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE An EST database from Sorghum: dark-grown seedlings
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordoniier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: JEN REV
 High quality sequence stop: 402
 POLYA-No.

FEATURES

source Location/Qualifiers
 1..493
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGI)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 104 a 147 c 179 g 63 t
 ORIGIN

alignment_scores:

Quality: 94.00 Length: 130
 Ratio: 1.106 Gaps: 7
 Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:

US-09-589-870-2 x BE357007/rev ..

Align seg 1/1 to reverse of: BE357007 from: 1 to: 493

```

10 AlValSerLeuThrValSerIleThrAlaSerAlaAspPr 26
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
459 TCTGCTCTGACCTCC...TCCGCTCTGCTGCGCGGCTCGGC 413
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
26 OserLyAspSerLySAlaGlnValSerAlaAlaGlyIleThrg 43
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
412 CGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
43 LThrTPTrpYrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
362 CCTGACCTGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 313
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
60 AspGlyAlaLeuThrGlyThrTyrgLuserAlaValGlyAsnAlaGluse 76
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
312 TCCGCTCTGCTTA...GCGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 266
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
76 rArgTyValLeuThrGlyArgTyraSpSerAlaProAlaThrAspGly. 92
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
265 GGCAGCGGGCTCTCTGACACCGACGAGCGCTCTCTGCTGCTG 216
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
93 ..SerGlyThrAlaLeuGlyThrPThrValAlaTrpLysAsnAsnTyraArg 108
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
215 TTTCCTCAACGGCGCGGCGGCTGAGACGCC.....TCTCCACATACC 175
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
109 AsnAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
174 ACCGAGCCTCTCTGACGCGGACGCGGACGCTGACCTGCGCGGT 125
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
120 rValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 133
   ::::|||||:|||||:|||||:|||||:|||||:|||||:
124 GGCCTATTGGTCA...GCTAGGGTG.....CTGTGGTTG 94

```

seq_name: gb_est70:BE359336

seq_documentation_block:

LOCUS BE359336 511 bp mRNA EST 20-JUL-2000
 DEFINITION DGI_40_E08.b1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
 sequence.

ACCESSION BE359336
 VERSION BE359336.1 GI:9300893

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.

REFERENCE 1 (bases 1 to 511)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 TITLE An EST database from Sorghum: dark-grown seedlings
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 490
 POLYA-No.

FEATURES
 source 1.511
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DG1)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector:
 Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
 made from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 107 a 154 c 166 g 64 t
 ORIGIN

alignment_scores:
 Quality: 94.00 Length: 130
 Ratio: 1.106 Gaps: 7
 Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:
 US-09-589-870-2 x BE359336/rev ..

Align seg 1/1 to reverse of: BE359336 from: 1 to: 511

```

10 AlAlaValSerLeuThrThrValSerIleThrAlaSerAlaSerAlaAspPr 26
   ::::::::::::::::::::|::::::::::::::::::|
460 TCTGCTTCCTTGACCTCC...TCCGCTCTGCTGCGCGCGGCTCGCGC 414

26 oSerLyAspSerLyAlaGlnValSerAlaAlaGlnAlaGlyIleThrg 43
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
413 CGCCTCTGCTTCCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 364

43 lYhTrpTyraSngInLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
363 CCTGACCTCGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314

60 AspGlyAlaLeuThrgIlyThrYrGluSerAlaValAlaGlyAlaGluSe 76
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
313 TCCGGCTCTCTTA...GGACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 267

76 rArGTYrValLeuThrgIlyArGTYrAspSerAlaProAlaThrAspGly 92
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
266 GCGCGCGCGGCTCTCTTCGACACCGACGACGAGGCTCTCTCTGGGTG 217

93 ..SerGlyThrAlaLeuGlyTrpThrValAlaTrpIlySAsnAsnTYrArg 108
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
216 TTTCCTCAACGGCGCGGCTGACACCC...TCTCCTCACTAC 176

109 AsnAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
175 ACCGGACCTCTCTTCGACGCGACGACGAGCTCGGACCTCGCGGT 126

120 rValGlyGlyAlaGlnAlaArGlyIleAsnThrgInTrpLeu 133
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
125 GGCCTATTGGTCA...GCTAGGGTG.....CTGTGGTTG 95

```

seq_name: gb_est70:BE361106

seq_documentation_block: 524 bp mRNA EST 20-JUL-2000
 LOCUS BE361106
 DEFINITION DGI.69.F08.b1.A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
 sequence.
 ACCESSION BE361106
 VERSION BE361106.1 GI:9302663
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 524)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 TITLE An EST database from Sorghum: dark-grown seedlings
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 520
 POLYA-No.

FEATURES
 source 1.524
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DG1)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector:
 Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
 made from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 107 a 156 c 195 g 66 t
 ORIGIN

alignment_scores:
 Quality: 94.00 Length: 130
 Ratio: 1.106 Gaps: 7
 Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:
 US-09-589-870-2 x BE361106/rev ..

Align seg 1/1 to reverse of: BE361106 from: 1 to: 524

```

10 AlAlaValSerLeuThrThrValSerIleThrAlaSerAlaSerAlaAspPr 26
   ::::::::::::::::::::|::::::::::::::::::|
456 TCTGCTTCCTTGACCTCC...TCCGCTCTGCTGCGCGCGGCTCGCGC 410

26 oSerLyAspSerLyAlaGlnValSerAlaAlaGlnAlaGlyIleThrg 43
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
409 CGCCTGCTTCCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

43 lYhTrpTyraSngInLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
359 CCTGACCTCGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310

60 AspGlyAlaLeuThrgIlyThrYrGluSerAlaValAlaGlyAlaGluSe 76
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
309 TCCGGCTCTCTTA...GGACCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 263

76 rArGTYrValLeuThrgIlyArGTYrAspSerAlaProAlaThrAspGly 92
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
262 GCGCGCGGCTCTCTTCGACACCGACGACGAGGCTCTCTCTGGGTG 213

```

```

93 ..SerglyThralaleuglytrpThralaIatrpIyAsnAsnTyArg 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212 TTCTCTACAGCGCGCGCGGTGACAGCC.....TCCTCCACTACC 172
109 AsnAlaHisSerAlaThr.....ThrTrpSerglyGlnTy 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 ACCGAGCCTCTCTGACGCGGACGCACTGCGGACCTGCGCGGT 122
120 rValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeu 133
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GGCCATTGTGTCA...GCTAGGTTG.....CTGTGTTG 91

seq_name: gb_est70:BE357620

seq_documentation_block: 538 bp mRNA EST 20-JUL-2000
LOCUS BE357620
DEFINITION DGI_21.C08.b1.A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE357620
VERSION BE357620.1 GI:9299177
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE 1 (bases 1 to 538)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@pratt.uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 534
POLYA-No.

FEATURES
source 1..538
location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced by mass excision."

BASE COUNT 109 a 161 c 203 g 65 t

ORIGIN
alignment_scores:
Quality: 94.00 Length: 130
Ratio: 1.106 Gaps: 7
Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:
US-09-589-870-2 x BE357620/rev ..

Align seq 1/1 to reverse of: BE357620 from: 1 to: 538

10 AlAValSerLeuThrThraValSerTleThraLaseRaAlaSerAlaAspPr 26
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 TCTGCTTCCTTGACCTCC...TCCGTCTCTGCTGCGCGCGGCTCTCGC 406

26 oSerLyAspSerLyAlaGluAlaArgIleAsnThrGlnTrpLeu 43
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

405 CGCCTCTGCTTCCTTGCTGTCTGTCTGCTCGCGCGGCTCAGCCAGC 356
43 LYThrTrpTyraSngInleuGlySerThrPheIleValThrAlaGlyAla 59
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 CCTGACCTCGACCTCTCTGTGTCTCAgTCTCGGACGAGAACGCTTCCGCC 306
60 AspGlyAlaLeuThrGlyThrTyrcIuSerAlaValGlyAlaGluSe 76
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305 TCCGCTCTCTTA...GCGACCTCGTCTCCGCGCGGAGCTGCTCTC 259
76 rArgTyValLeuThrGlyArgTyraSeraAlaProAlaThrAspGly 92
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 GCGCGCGCGCTCTCTTCTGACACCGGACGAGAGCCCTCTGCTGGTG 209
93 ..SerglyThralaleuglytrpThralaIatrpIyAsnAsnTyArg 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 TTCTCTACAGCGCGCGGTGACAGCC.....TCCTCCACTACC 168
109 AsnAlaHisSerAlaThr.....ThrTrpSerglyGlnTy 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 ACCGAGCCTCTCTGACGCGGACGCACTGCGGACCTGCGCGGT 118
120 rValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeu 133
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 GGCCATTGTGTCA...GCTAGGTTG.....CTGTGTTG 87

seq_name: gb_est70:BE357934

seq_documentation_block: 558 bp mRNA EST 20-JUL-2000
LOCUS BE357934
DEFINITION DGI_23.D10.b1.A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE357934
VERSION BE357934.1 GI:9299491
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE 1 (bases 1 to 558)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@pratt.uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 528
POLYA-No.

FEATURES
source 1..558
location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced by mass excision."

BASE COUNT 115 a 169 c 208 g 66 t

ORIGIN
alignment_scores:
Quality: 94.00 Length: 130
Ratio: 1.106 Gaps: 7

```

